

DRINKING WATER IN RURAL SASKATCHEWAN: PUBLIC PERCEPTION  
OF WATER QUALITY AND HEALTH RISKS, AND DIRECT AND INDIRECT EFFECTS  
OF DRINKING WATER QUALITY ON CHRONIC DISEASE

A Thesis Submitted to the College of  
Graduate and Postdoctoral Studies  
In Partial Fulfillment of the Requirements  
For the Degree of Doctor of Philosophy  
In the Department of Large Animal Clinical Sciences  
University of Saskatchewan  
Saskatoon

By  
Lianne M. McLeod

©Lianne Mary McLeod, February, 2017. All rights reserved.

## **PERMISSION TO USE**

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Large Animal Clinical Sciences Department  
Western College of Veterinary Medicine  
University of Saskatchewan  
52 Campus Drive  
Saskatoon, Saskatchewan, S7N 5B4, Canada

OR

Dean  
College of Graduate Studies and Research  
University of Saskatchewan  
107 Administration Place  
Saskatoon, Saskatchewan, S7N 5A2 Canada

## **ABSTRACT**

In rural areas of Saskatchewan, tap water supplied to houses is not typically tested for quality and safety with the same intensity as public supplies that serve larger populations. Consequently, rural residents might be at greater risk of exposure to poor quality water and any resulting health impacts. The overall objective of this study was to investigate if poor quality water in rural areas was directly or indirectly associated with increased occurrence of diabetes and cardiovascular disease. Perceptions about water quality and safety and how these relate to choices about drinking water among rural Saskatchewan residents were also investigated. Existing water surveillance data from both public water supplies and private wells in SK were summarized using a combination of principal components analysis and geostatistics. The summarized water data were used to estimate regional water quality exposure indicators for use in Bayesian hierarchical models examining ecological associations with health outcomes derived from administrative health data.

A quarter of 2065 respondents to a questionnaire sent to rural SK residents reported being unsatisfied with the aesthetic quality of their water, although fewer (12%) believed their water was unsafe to drink. Of the respondents, 31% reported drinking primarily bottled water, while 61% reported drinking tap water at least daily and 48% reported treating their water in the home. The type of water supply along with past experiences and familiarity with the water were consistently associated with risk perception. As expected, perception of quality and risk were important predictors of drinking water choices; aspects of familiarity, experience, and type of water supply were also important.

The parameters listed under health standards and aesthetic objectives grouped differently for public water supply and private well data following the application of principal components analysis, suggesting residents using different types of water supplies may be exposed to different patterns of contaminants. Summarizing water quality data through geostatistical models resulted in attenuation of extreme concentrations recorded in the observed data, but appeared to predict trends in water quality that could be useful for prioritizing monitoring efforts and public health messaging about water testing for private wells.

Overall, poor groundwater quality was not associated with increased occurrence of diabetes or cardiovascular disease. An increase in principal component scores for public water supplies, characterized mainly by the presence of high levels of hardness and magnesium, was associated with a decrease in the prevalence of ischemic heart disease. This finding was consistent with previously reported results in other regions, and raises the question of whether the in-home treatment of water to remove high mineral content could inadvertently increase the risk of cardiovascular disease. Studies with individual-level exposure measures are recommended to more definitively characterize potential associations between water quality and chronic disease.

This study used innovative methods to address gaps in knowledge about perceptions of water quality and risk and drinking water choices for people living in rural SK, summarized water quality over a large region of the province, and investigated associations between water quality and the occurrence of important chronic diseases in rural Saskatchewan.

## **ACKNOWLEDGEMENTS**

I would like to thank my graduate student committee, Dr. Cheryl Waldner, Dr. Tasha Epp, Dr. Lalita Bharadwaj and Dr. Cindy Feng, for their guidance and unwavering support. I am especially grateful to Cheryl Waldner, whose patience appears to know no bounds. I thank her for giving me the latitude to work through problems at my own pace, offering encouragement during the difficult bits, and for supporting my involvement in side projects along the way. I consider myself extremely fortunate that she agreed to supervise me.

I am grateful to the Water Security Agency, especially Sam Ferris and Lorelei Ford, for providing access to the data that made this thesis work possible. Special thanks to Lorelei Ford for going the extra mile to ensure we could access the private well data, and especially for her advice, support, and friendship.

I am also indebted to the entire team at the Health Quality Council for their assistance in accessing the health data. In particular, many thanks go to Shan Jin for her guidance and her work in extracting my study cohorts.

I am very grateful for the generous financial support that made this research possible, especially the Saskatchewan Health Research Foundation and the Interprovincial Graduate Student Fellowship. Additional support from the NSERC/Create ITraP program, Saskatchewan Innovation and Opportunity Scholarship and ESRI Canada Scholarship is also gratefully acknowledged.

I would also like to thank all of my fellow students who made this journey easier. In particular, I would like to thank Silvia and Nicolas, whose friendship, kindness, and support will always be treasured. Also Jasmine, whose encouragement, especially during the writing phase, will never be forgotten. Thanks also go to Jackie Gabriel, Saira Bilal and Rob Bechard for working on data entry for the questionnaire portion of this research.

Last but not least, I would like to thank my family for their unwavering and unconditional support, especially Kenny and Chris.

## **DEDICATION**

This thesis is dedicated to Kenny and Chris. It could not have been easy for you during the times I had so much to do. Nevertheless, you both grew into fine young men while I worked on this, and I am so grateful that you are my sons.

## TABLE OF CONTENTS

PERMISSION TO USE .....	i
ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iv
DEDICATION .....	v
TABLE OF CONTENTS .....	vi
LIST OF TABLES .....	xii
LIST OF FIGURES .....	xx
LIST OF ABBREVIATIONS .....	xxiii
CHAPTER 1: INTRODUCTION: DRINKING WATER IN RURAL SASKATCHEWAN, PUBLIC PERCEPTION OF WATER QUALITY AND HEALTH RISKS, AND DIRECT AND INDIRECT EFFECTS OF DRINKING WATER QUALITY ON CHRONIC DISEASE. ....	1
1.1 Drinking Water in Saskatchewan: Guidelines and Regulation .....	2
1.2 Overview of Drinking Water Sources in Rural and Remote Areas of Saskatchewan .....	4
1.3 Vulnerability of rural populations to water-related health .....	5
1.4 Perceptions of Risk and Quality and Impacts on Consumption .....	7
1.5 Associations between arsenic in drinking water and chronic disease .....	10
1.5.1. Associations between arsenic and type 2 diabetes mellitus .....	10
1.5.2 Associations between arsenic and cardiovascular disease .....	13
1.5.3 Effect modifiers on associations between arsenic and chronic disease .....	16
1.6 Indirect effects of Poor Quality Drinking Water on Diabetes and Cardiovascular Disease .....	17
1.7. Assessing effects of mixtures of contaminants in drinking water .....	18
1.8 Objectives of research .....	18
1.8.1 Investigate factors influencing perceptions of water quality and safety among rural Saskatchewan residents .....	19

1.8.2 Factors affecting choices to drink tap water or bottled water and to treat household tap water	20
1.8.3 Summarize existing surveillance data about drinking water quality.....	20
1.8.4 Investigate associations between water quality and type 2 diabetes mellitus .....	21
1.8.5 Investigate associations between water quality and cardiovascular disease .....	22
1.9 References .....	23
 CHAPTER 2: RISK FACTORS ASSOCIATED WITH PERCEPTIONS OF DRINKING WATER QUALITY IN RURAL SASKATCHEWAN.....	
	31
2.1 Abstract .....	32
2.2 Introduction .....	32
2.3 Materials and methods .....	34
2.3.1 Design .....	34
2.3.2 Outcomes and risk factors.....	35
2.3.3 Statistical analysis.....	37
2.4 Results .....	39
2.4.1 Having any aesthetic complaint .....	41
2.4.2 The perception that tap water is not safe to drink .....	45
2.4.3 Fear that water will become contaminated.....	47
2.4.4 Perception that someone became ill from drinking the household's tap water.....	48
2.4.5 Summary of risk factors.....	48
2.5 Discussion .....	49
2.5.1 Aesthetic complaints .....	50
2.5.2 Perception that tap water is not safe.....	51
2.5.3 Fear that water will become contaminated.....	52
2.5.4 Illness due to drinking tap water .....	53
2.5.5 Limitations .....	53
2.6 Conclusions .....	54
2.7 References .....	55



CHAPTER 3: RISK FACTORS ASSOCIATED WITH THE CHOICE TO DRINK BOTTLED WATER AND TAP WATER IN RURAL SASKATCHEWAN .....	57
3.1 Abstract .....	58
3.2 Introduction .....	58
3.3 Materials and Methods .....	59
3.3.1. Design .....	59
3.3.2 Outcomes and Potential Risk Factors of Interest .....	60
3.3.3 Statistical Analysis .....	62
3.4 Results .....	64
3.4.1 Descriptive Statistics.....	64
3.4.2 Choosing Primarily Bottled Water.....	66
3.4.3 Consuming Tap Water Daily .....	70
3.4.4 In-home Treatment of Tap Water .....	73
3.5 Discussion .....	76
3.5.1 Consuming Primarily Bottled Water.....	76
3.5.2 Consuming Tap Water Daily .....	79
3.5.3 Treating Tap Water .....	81
3.5.4 Limitations .....	82
3.6 Conclusions .....	82
3.7 References .....	84
CHAPTER 4: EVALUATION OF THE USE OF EMPIRICAL BAYESIAN KRIGING FOR SUMMARIZING GROUNDWATER ARSENIC CONCENTRATIONS AND THE USE PRINCIPAL COMPONENTS ANALYSIS COMBINED WITH EMPIRICAL BAYESIAN KRIGING TO SUMMARIZE MIXTURES OF TRACE METALS AND MAJOR IONS IN GROUNDWATER IN SOUTHERN SASKATCHEWAN.....	86
4.1 Abstract .....	87
4.2 Introduction .....	88

4.3 Materials and Methods .....	91
4.3.1 Study Area .....	91
4.3.2 Data Sources .....	92
4.3.3 Principal Components Analysis .....	94
4.3.4 Geostatistical Analysis .....	95
4.4 Results .....	97
4.4.1 Principal Components Analysis .....	101
4.4.2 Geostatistical Analysis .....	104
4.5 Discussion .....	112
4.5.1 Principal Components Analysis .....	113
4.5.2 Geostatistical Analysis .....	114
4.5.3 Limitations .....	118
4.6 Conclusion .....	120
4.7 References .....	121
CHAPTER 5: ECOLOGICAL ANALYSIS OF ASSOCIATIONS BETWEEN GROUNDWATER QUALITY AND TYPE 2 DIABETES INCIDENCE AND PREVALENCE IN RURAL SASKATCHEWAN USING BAYESIAN HIERARCHICAL MODELS .....	125
5.1 Abstract .....	126
5.2 Introduction .....	127
5.3 Materials and Methods .....	129
5.3.1 Study Area .....	129
5.3.2 Geographic Units for Analysis .....	130
5.3.3 Diabetes Data .....	131
5.3.4 Water Data .....	133
5.3.5 Covariates .....	135
5.3.6 Bayesian Statistical Analyses of Associations between Water Quality Data and Diabetes .....	136
5.3.7 Frequentist Statistical Analyses of Associations between Water Quality Data and Diabetes .....	139

5.3.8 Model Results and Assessment of Fit .....	139
5.4 Results .....	141
5.4.1 Associations between Arsenic Concentration and Diabetes Incidence.....	147
5.4.2 Associations between Arsenic Concentration and Diabetes Prevalence.....	150
5.4.3 Associations between Health Standards and Diabetes Incidence .....	152
5.4.4 Associations between Health Standards and Diabetes Prevalence .....	155
5.4.5 Associations between Aesthetic Objective Factor Scores and Diabetes Incidence .....	159
5.4.6 Associations between Aesthetic Objective Factor Scores and Diabetes Prevalence .....	163
5.4.7 Summary of results .....	166
5.4.8 Residuals and Model Fit .....	168
5.4.9 Prior sensitivity .....	169
5.5 Discussion .....	170
5.6 Conclusions .....	179
5.7 References .....	180
CHAPTER 6: ECOLOGICAL ANALYSIS OF ASSOCIATIONS BETWEEN GROUNDWATER QUALITY AND HYPERTENSION AND CARDIOVASCULAR DISEASE IN RURAL SASKATCHEWAN USING BAYESIAN HIERARCHICAL MODELS .....	184
6.1 Abstract .....	185
6.2 Introduction .....	186
6.3 Materials and Methods .....	188
6.3.1 Study Area .....	188
6.3.2 Geographic Units for Analysis.....	189
6.3.3 Disease Data.....	190
6.3.4 Water Data .....	193
6.3.5 Covariates .....	195
6.3.6 Bayesian Statistical Analyses of Associations between Water Quality Data and Outcomes ..	196
6.3.7 Frequentist Statistical Analyses of Associations between Water Quality Data and Outcomes	198

6.3.8 Model Results and Assessment of Fit .....	198
6.4 Results .....	200
6.4.1 Hypertension and arsenic concentrations.....	206
6.4.2 Hypertension and health standards PC scores.....	209
6.4.3 Hypertension and aesthetic objective PC scores.....	213
6.4.4 Ischemic heart disease and arsenic concentrations .....	217
6.4.5 Ischemic heart disease and health standards PC scores .....	219
6.4.6 Ischemic heart disease and aesthetic objective PC scores .....	223
6.4.7 Stroke and arsenic concentrations.....	227
6.4.8 Stroke and health standards PC scores.....	229
6.4.9 Stroke and aesthetic objective PC scores.....	233
6.4.10 Summary of results .....	237
6.4.11 Evaluation of model fit .....	240
6.4.12 Prior sensitivity .....	241
6.5 Discussion .....	242
6.6 Conclusions .....	250
6.7 References.....	252
CHAPTER 7: CONCLUSIONS .....	258
7.1 Strengths of the Research.....	263
7.2 Limitations .....	264
7.3 Future research .....	266
7.4 Summary .....	268
7.5 References .....	269
APPENDIX.....	270

## LIST OF TABLES

Table 2.1 Total numbers of complete and missing responses and the proportion of respondents at each level for each of the outcomes modeled and risk factors evaluated .....	40
Table 2.2 Comparison of key demographic variables between the sample population and the Statistics Canada 2011 Census of Population for the rural Census Subdivisions included within the survey regions. ....	41
Table 2.3 Proportion of respondents identifying each type of aesthetic complaint overall and by the type of water source used.....	42
Table 2.4 Risk factors significantly associated with having any type of aesthetic complaint with the household's tap water in the final multivariable model.....	43
Table 2.5 Risk factors significantly associated, in the final multivariable model, with the perception that household tap water is not safe to drink.....	46
Table 2.6 Summary of the analyses for all water quality or risk perception outcomes including the number of observations used and the risk factors associated with the final model for each outcome (effect modifications italicized). ....	49
Table 3.1 Number of complete and missing responses and the proportion of respondents at each level for the outcomes modeled and risk factors evaluated. ....	65
Table 3.2 Comparison of frequency of key demographic variables in the survey sample population and the Statistics Canada 2011 Census of Population for the rural Census Subdivisions included within the survey regions.....	66
Table 3.3 Risk factors associated with the primary consumption of bottled water in the final multivariable model. ....	68
Table 3.4 Risk factors included in final multivariable model for daily consumption of tap water. ....	71
Table 3.5 Risk factors included in final multivariable model for in-home treatment of tap water. ....	74

Table 3.6 Summary of the analyses for consuming primarily bottled water, consuming tap water daily and choosing to treat tap water, including the number of observations used and the risk factors associated with the final model for each outcome (effect modifications italicized).....	75
Table 4.1 Descriptive statistics for individual samples in public water supplies in Saskatchewan from 1985-2012, along with the Saskatchewan standards and objectives, percent of samples exceeding the respective standard or objective, and percent of samples below detection limit for the water parameters included in the analysis.....	98
Table 4.2 Descriptive statistics for individual samples in private water supplies in Saskatchewan from 1996-2011, along with the Saskatchewan standards and objectives, percent of samples exceeding the respective standard or objective, and percent of samples below detection limit for the water parameters included in the analysis.....	99
Table 4.3 Descriptive statistics for predicted concentrations after repeated measures modeling for each parameter by site for public water supplies in Saskatchewan, along with Saskatchewan standards and objectives, and the percent of sites with a predicted concentration exceeding the standard or objective. Also shown is a summary of the number of samples per site incorporated in repeated measures models for each parameter. ....	100
Table 4.4 Summary of test statistics to assess the adequacy of sampling and correlation among variables in each of the data sets for PCA. ....	101
Table 4.5 Principal components analysis on public water supplies and private well water supplies: varimax rotated principal components patterns, and eigenvalues and percent of variance explained for each retained component. The loadings in bold font are the maximum loading for each variable. ....	103
Table 4.6 Summary of variogram models for each variable in public and private water supplies. Inputs into models were lag distance (Average nearest neighbor distance was used) and number of lags. Large scale trends identified in the data are reported, along with results of variogram modeling (model form, Moran's I, Geary's C, nugget, and range).....	105
Table 4.7 Values of Lin's concordance correlation coefficient comparing values in the validation datasets and values predicted by ordinary kriging, universal kriging, and empirical Bayesian	

kriging for log arsenic concentration and component scores for public and private water supplies. .....	106
Table 4.8 Values of Spearman’s rank coefficient for correlation between actual values and values predicted by ordinary kriging, universal kriging, and empirical Bayesian kriging for log arsenic concentration and component scores for public and private water supplies.....	107
Table 4.9 Values of the kappa statistic for agreement between values predicted by ordinary, universal and empirical Bayesian kriging and measured values being over the 75 <sup>th</sup> percentile for observed arsenic concentrations and component scores for public and private water supplies. Also included are kappa statistics for being over the drinking water standard (0.01 mg/L) for arsenic, and sensitivity of the various kriging techniques for identifying predicted values over the 75 <sup>th</sup> percentile of the observed data. ....	109
Table 5.1 Case definitions and International Classification of Diseases (ICD) codes used to identify cases of diabetes in Saskatchewan administrative health data. ....	132
Table 5.2 Case definitions and International Classification of Diseases codes used to identify cases of Chronic Obstructive Pulmonary Disease in Saskatchewan administrative health data. ....	135
Table 5.3 Number of rural Saskatchewan residents included in the study by year and by age group. ....	143
Table 5.4 Diabetes cumulative incidence and prevalence in each time period analyzed for the population of the study area in southern Saskatchewan ages 35 to 75 who did not live in a city. .....	144
Table 5.5 Summary of principal components analysis of groundwater parameters used as risk factors in models, showing component loadings along with the eigenvalue and cumulative variances for each retained component. The maximum loading for each parameter is indicated in bold. Details of the PCA analysis can be found in Chapter 4.....	145
Table 5.6 Median and 5 <sup>th</sup> and 95 <sup>th</sup> percentile values for the estimated mean exposures for principal component scores for each geographic and the corresponding values for individual water supplies used as input for kriging for public water supplies and private wells. ....	146

Table 5.7 Comparison of Bayesian and frequentist models of the association between arsenic concentrations in drinking water and cumulative diabetes incidence 2010-2012 in rural southern Saskatchewan.....	148
Table 5.8 Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and 3-year cumulative incidence of diabetes for the time periods 2006-2008 through 2009-2011 in rural southern Saskatchewan.....	149
Table 5.9 Comparison of Bayesian and frequentist models of the association between arsenic concentrations in drinking water and diabetes prevalence for 2010 in rural southern Saskatchewan.....	150
Table 5.10 Summary of the results for frequentist GLMMs examining the association between arsenic concentrations in drinking water and diabetes prevalence for 2006-2009 in rural southern Saskatchewan.....	151
Table 5.11 Comparison of Bayesian and frequentist models of the association between drinking water health standards principal component scores and cumulative diabetes incidence 2010-2012 in rural southern Saskatchewan. ....	152
Table 5.12 Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and 3-year cumulative incidence of diabetes for the time periods 2006-2008 through 2009-2011 in rural southern Saskatchewan..	154
Table 5.13 Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and diabetes prevalence for 2010 in rural southern Saskatchewan. ....	156
Table 5.14 Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and diabetes prevalence for 2006-2009 in rural southern Saskatchewan. ....	158
Table 5.15 Comparison of Bayesian and frequentist models of the association between drinking water aesthetic objectives principal component scores and diabetes cumulative incidence 2010-2012 in rural southern Saskatchewan. ....	160
Table 5.16 Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and 3-year cumulative incidence	



of diabetes for the time periods 2006-2008 through 2009-2011 in rural southern Saskatchewan.	162
Table 5.17 Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and diabetes prevalence for 2010 in rural southern Saskatchewan.	164
Table 5.18 Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and diabetes prevalence for 2006-2009 in rural southern Saskatchewan.	165
Table 5.19 Summary of effect estimates from Bayesian analysis of associations between drinking water risk factors from public water supply and private well data and cumulative incidence of diabetes for 2010-2012 and prevalence for 2010. Results in bold type indicate that the 95% credible interval for that effect estimate did not include 1.	166
Table 5.20 Total number of non-significant results and positive and negative associations between the water quality risk factors and diabetes incidence or prevalence from the frequentist GLMM analysis over 5 time periods.	167
Table 5.21 Values for Moran's I for global spatial autocorrelation for residuals from frequentist models for diabetes incidence and prevalence from public and private water supplies. Risk factors include arsenic concentrations, health standards PC scores, or aesthetic objective PC scores as risk factors. Each model was adjusted for prevalence of COPD, education level and income. Residuals for null models or models for diabetes incidence and prevalence with no fixed effects were also analyzed.	168
Table 5.22 Results of prior sensitivity analysis comparing the reported Bayesian model of associations between arsenic concentrations in private wells categorized by quintiles and diabetes incidence in 2010 to models with two alternative hyperprior specifications for the unstructured and structured random effects.	170
Table 6.1 Case definitions and International Classification of Diseases (ICD) codes used to identify cases of hypertension, ischemic heart disease, and stroke, as well as chronic obstructive pulmonary disease for use as a covariate, in Saskatchewan administrative health data.	192

Table 6.2 Hypertension, ischemic heart disease, and stroke prevalence in the study area population ages 35 to 75 for those not living in a Saskatchewan city. ....	202
Table 6.3 Summary of principal components analysis of groundwater parameters for health standards and aesthetic objectives from public water supplies and private wells. The loadings for each component are shown along with the eigenvalue and cumulative variance explained for each retained component; the maximum loading for each parameter is indicated in bold. Details of the PCA analysis can be found in Chapter 4. ....	204
Table 6.4 Median, 5 <sup>th</sup> and 95 <sup>th</sup> percentile values for the mean exposures for principal component scores estimated for each geographic unit (left) and the scores for individual water supplies used as input for kriging for public water supplies and private wells. ....	205
Table 6.5 Comparison of Bayesian and frequentist models for the association between arsenic concentrations in drinking water and prevalence of hypertension for 2010 in rural southern Saskatchewan. ....	207
Table 6.6 Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and prevalence of hypertension for 2006-2009 in rural southern Saskatchewan. ....	208
Table 6.7 Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and hypertension prevalence for 2010 in rural southern Saskatchewan. ....	210
Table 6.8 Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and hypertension prevalence for 2006-2009 in rural southern Saskatchewan. ....	212
Table 6.9 Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and hypertension prevalence for 2010 in rural southern Saskatchewan. ....	214
Table 6.10 Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and hypertension prevalence for 2006-2009 in rural southern Saskatchewan. ....	216

Table 6.11 Comparison of Bayesian and frequentist models for the association between arsenic concentrations in drinking water and ischemic heart disease prevalence for 2010 in rural southern Saskatchewan.....	217
Table 6.12 Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and ischemic heart disease prevalence for 2006-2009 in rural southern Saskatchewan.....	218
Table 6.13 Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and ischemic heart disease prevalence for 2010 in rural southern Saskatchewan. ....	220
Table 6.14 Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and ischemic heart disease prevalence for 2006-2009 in rural southern Saskatchewan. ....	222
Table 6.15 Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and ischemic heart disease prevalence for 2010 in rural southern Saskatchewan. ....	224
Table 6.16 Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and ischemic heart disease prevalence for 2006-2009 in rural southern Saskatchewan. ....	226
Table 6.17 Comparison of Bayesian and frequentist models for the association between arsenic concentrations in drinking water and stroke prevalence for 2010 in rural southern Saskatchewan. ....	227
Table 6.18 Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and stroke prevalence for 2006-2009 in rural southern Saskatchewan.....	228
Table 6.19 Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and stroke prevalence for 2010 in rural southern Saskatchewan. ....	230

Table 6.20 Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and stroke prevalence for 2006-2009 in rural southern Saskatchewan. ....	232
Table 6.21 Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and stroke prevalence for 2010 in rural southern Saskatchewan. ....	234
Table 6.22 Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and stroke prevalence for 2006-2009 in rural southern Saskatchewan. ....	236
Table 6.23 Summary of effect estimates from Bayesian analysis of associations between drinking water risk factors from public water supply and private well data and hypertension, ischemic heart disease, and stroke prevalence for 2010. Effect estimates for which the 95% credible interval did not include 1 are indicated with bold type.....	238
Table 6.24 Total number of non-significant results and positive and negative associations between the water quality risk factors and hypertension, ischemic heart disease, and stroke prevalence from the frequentist GLMM analysis over 5 time periods. ....	239
Table 6.25 Values for Moran's I for global spatial autocorrelation for residuals from frequentist models for the association between measures of water quality and cardiovascular disease from public and private water supplies.....	241
Table 6.26 Results of prior sensitivity analysis comparing the reported Bayesian model of associations between health standards principal components in public water supplies and stroke prevalence in 2010 to models with two alternative hyperprior specifications for the precision of the unstructured and structured random effects. ....	242

## LIST OF FIGURES

Figure 2.1 Causal diagram used to guide model development for each of the outcomes.....	36
Figure 2.2 Predicted probabilities of having any aesthetic complaint, comparing those who had a drinking water advisory to those who did not, for men and women in each age category averaged over the use of water delivered by truck and length of time residing in community. CI: confidence interval.....	44
Figure 2.3 Predicted probabilities of believing the tap water is not safe to drink for the interaction between reporting a drinking water advisory and use of a groundwater source, separated by whether or not an aesthetic complaint about the water was identified, and averaged over the use of a community water supply. CI: confidence interval.....	47
Figure 3.1 Generalized causal diagram used to direct model development for each of the outcomes related to water consumption and treatment choices (potential interactions not diagrammed for simplicity).....	61
Figure 3.2 Predicted probability of consuming primarily bottled water by presence of aesthetic complaint and the belief that tap water is not safe to drink, separated by whether or not household had a water advisory in the past averaged over all length of time residing in area, whether a community water supply is used, and in-home treatment of tap water.....	69
Figure 3.3 Predicted probability of consuming primarily bottled water by use of a community water supply and in-home treatment of tap water separated by length of residence in the area averaged over reported aesthetic complaints, agreement that tap water is not safe to drink and whether the household experienced a water advisory.....	69
Figure 3.4 Predicted probability of consuming tap water daily by presence of aesthetic complaint and the belief that tap water is not safe to drink, separated by in-home treatment of tap water averaged over all whether household had experienced a water advisory and length of time resided in area. ....	72
Figure 3.5 Predicted probability of treating tap water in-home by use of a private water supply and the belief that tap water is not safe to drink, separated by whether or not children reside in household, adjusted for whether the home was in a town and whether a community water supply was used.....	74

Figure 4.1 The study area (solid line) within the province of Saskatchewan along with the density of sampling sites for public water supplies (left) and private wells (right) across the province. Inset map shows location of Saskatchewan within Canada.....	92
Figure 4.2 Limits of agreement plots for between observed and predicted $\ln(\text{arsenic concentrations in mg/L})$ for the different types of kriging for public supplies (top row) and private supplies (bottom row). .....	110
Figure 4.3 Arsenic concentrations ( $\mu\text{g/L}$ ) predicted by empirical Bayesian kriging for public water supply data (left) and private well data (right) in study area in southern Saskatchewan..	110
Figure 4.4 Empirical Bayesian kriging prediction maps for natural logarithm transformed arsenic concentrations (a) along with the prediction standard error maps (b) for public supplies (on left) and private supplies (on right) in study area in southern Saskatchewan. ....	111
Figure 5.1 A map showing the province of Saskatchewan, the boundary of the study area for evaluation of associations between water quality and diabetes, and rural municipalities within the study area. Inset map shows the location of Saskatchewan within Canada.....	130
Figure 5.2 Flow chart showing identification and aggregation of incident cases for the cumulative three-year incidence case counts used for modeling, and identification of prevalent cases. ....	133
Figure 5.3 Summary of the process and equations used to obtain stratum-specific expected case counts for diabetes incidence and prevalence for each geographic unit. ....	137
Figure 5.4 A flow chart outlining process of evaluating potential associations between water quality risk factors and diabetes outcomes in Bayesian and frequentist models. ....	141
Figure 5.5 A map of Saskatchewan showing geographic regions for analysis of the association between measures of water quality and diabetes developed by amalgamation of RMs. ....	142
Figure 6.1 A map showing the province of Saskatchewan and the study area for the analysis of associations between water quality and hypertension, IHD and stroke along with the rural municipalities within the study area prior to aggregation. The small inset map provides context regarding the location of Saskatchewan within Canada. ....	189

Figure 6.2 A flow chart outlining process of evaluating potential associations between water quality risk factors and hypertension, ischemic heart disease and stroke in Bayesian and frequentist models. ....	200
Figure 6.3 A map of Saskatchewan showing geographic regions formed by amalgamation of RMs for analysis of the association between measures of water quality and hypertension, ischemic heart disease, and stroke. ....	201

## LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
CCC	Concordance Correlation Coefficient
CCS	Consolidated Census Subdivision
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CrI	Credible Interval
CVD	Cardiovascular Disease
DIC	Deviance Information Criterion
GCDWQ	Guidelines for Canadian Drinking Water Quality
GLMM	Generalized Linear Mixed Model
ICD	International Classification of Disease
IHD	Ischemic Heart Disease
KMO	Kaiser-Meyer-Olkin Measure of Sampling Adequacy
OR	Odds Ratio
P5	5 <sup>th</sup> Percentile
P95	95 <sup>th</sup> Percentile
PC	Principal Components
pc	Postal Code
PCA	Principal Components Analysis
Q1	1 <sup>st</sup> Quartile
Q3	3 <sup>rd</sup> Quartile
RE	Random Effect
RM	Rural Municipality
RR	Risk Ratio
RWQAP	Rural Water Quality Advisory Program
SD	Standard Deviation
SE	Standard Error
SK	Saskatchewan
SMR	Standardized Mortality Ratio
VPC	Variance Partition Coefficient
WSA	Water Security Agency



**CHAPTER 1: INTRODUCTION: DRINKING WATER IN RURAL SASKATCHEWAN,  
PUBLIC PERCEPTION OF WATER QUALITY AND HEALTH RISKS, AND DIRECT  
AND INDIRECT EFFECTS OF DRINKING WATER QUALITY ON CHRONIC  
DISEASE.**

In urban areas, most Canadians enjoy a plentiful supply of water, which is regularly tested to meet drinking water standards and a level of service not shared by many rural and remote areas of the country. In Saskatchewan, private water supplies are not regulated, and residents with private water supplies are solely responsible for ensuring the safety of their water supply. Without routine testing, rural Saskatchewan residents could be more vulnerable to water-related health issues. In addition, public supplies serving small populations are typically not tested with the same intensity as those that serve larger populations. As a consequence, residents in rural areas of Saskatchewan, which make up a third of the Saskatchewan population (Statistics Canada, 2011), may be at increased risk of exposure to drinking water supplies that may not meet reported standards for health or objectives that impact palatability.

### **1.1 Drinking Water in Saskatchewan: Guidelines and Regulation**

In Canada, regulation of drinking water falls under provincial jurisdiction and each province establishes regulations pertaining to drinking water supplies. In Saskatchewan, the oversight of drinking water depends on the type of supply and is a three-tiered system. Public water supplies are defined as those that feed a distribution system and have a flow rate of greater than 18 cubic meters per day. These supplies are monitored by the Saskatchewan Water Security Agency (WSA) under The Water Regulations (Government of Saskatchewan, 2002a) and include municipal waterworks as well as waterworks in provincial parks, other distribution systems such as those found in trailer courts, institutions, and Hutterite Colonies, as well as some rural water pipelines. The WSA was formed in 2012 to bring together several branches responsible for water management in the province; prior to this the Ministry of Environment was responsible for public water supplies.

Semi-public water supplies are monitored by the Ministry of Health under the Health Hazard Regulations (Government of Saskatchewan, 2002b). These systems are those that serve the public but have a flow rate of under 18 cubic meters per day and lack a distribution network or serve at least three but less than 15 service connections.

Private water supplies serve a single residence or farm. Private water supplies are not monitored or regulated with the exception of requirements pertaining to construction and abandonment under The Ground Water Regulations (Government of Saskatchewan, 1966). Therefore, the

landowner bears sole responsibility for ensuring the quality and safety of their water supply. While advice regarding testing of private supplies is available, the cost as well as collection and transport of samples are typically the responsibility of the landowner.

Landowners who rely on private water supplies benefitted for several years from the WSA's Rural Water Quality Advisory Program (RWQAP), a program recognized for its innovation (Charrois, 2010). The RWQAP provided support and advice regarding water supply management and testing as well as providing a subsidized price for testing for a wide range of potential contaminants. As a result of this program, the WSA now has a database of over 5000 samples from private wells across Saskatchewan. However, the RWQAP was discontinued in 2011.

Landowners with private water supplies may lack the knowledge and resources to regularly test their household water for a wide range of contaminants. Barriers to adequate testing in rural areas include cost and inconvenience (Charrois, 2010). With respect to chemical contaminants, semi public supplies are required to test for major ions on an annual basis, although regional Health Authorities may require more frequent testing at their discretion (Government of Saskatchewan, 2002b). For public supplies, the requirements for testing, including the list of parameters and sampling frequency, are based on the size of the population served (Water Security Agency, 2016). Consequently, smaller public supplies are tested at a reduced intensity compared to supplies serving larger population centres. Waterworks in smaller centers may also suffer from a lack of resources, including highly trained personnel, when compared to water treatment facilities in urban centers. As a result, residents in remote and rural areas, whether they use a private water supply or access a smaller public water supply, may be less likely to have high-quality, regularly tested water at their household tap. Therefore, compared to urban residents, residents of rural areas could be at greater risk of exposure to drinking water of unknown or poor quality, possibly with elevated concentrations of contaminants posing a hazard to human health.

A set of non-binding guidelines are established by the Federal Provincial Territorial Committee on Drinking Water, and published by Health Canada (Health Canada, 2014). The Guidelines for Canadian Drinking Water Quality (GCDWQ) are periodically reviewed and updated if necessary, based on current published scientific literature. The GCDWQ are categorized as

health-based guidelines for contaminants that have been identified as potentially causing adverse effects on human health. Aesthetic guidelines refer to contaminants that affect palatability of drinking water but are not known to cause adverse health outcomes. Operational guidelines cover contaminants that might interfere with effective water treatment or cause problems in water infrastructure such as erosion of plumbing. The GCDWQ serves as a resource and guidance document for the establishment of guidelines, regulations, and drinking water-related legislations by each of the province and territories.

Saskatchewan's Drinking Water Quality Standards and Objectives (Water Security Agency, n.d.) were developed for regulated water supplies. Standards are legally enforceable for regulated water supplies and refer to contaminants that could adversely affect human health. Standards are described in the Water Regulations 2002 (Government of Saskatchewan, 2002a) and include Bacteriological Standards, Turbidity Standards, and contaminants that fall under categories of Chemical-Health Standards, Pesticides, and Radiological Standards. Objectives refer to contaminants or characteristics that adversely affect the aesthetic quality (e.g., taste, odor, color) of drinking water and affect consumer acceptance of water for drinking or hygienic purposes. It is not mandatory for regulated water supplies to meet the objective guidelines because they are not considered to routinely constitute health risks, with the caveat that they "may represent a health risk to some people if found in excessive concentrations" (Water Security Agency, n.d.).

Several exemptions exist to the provincial jurisdiction over drinking water regulations. First Nations Reserves, and federal lands such as national parks as well as federal institutions such as prisons fall under the jurisdiction of the federal government and are not subject to provincial drinking water regulations.

## **1.2 Overview of Drinking Water Sources in Rural and Remote Areas of Saskatchewan**

Approximately 85% of Saskatchewan's population access water from public water supplies for domestic use, while the remaining 15% rely on privately owned water supplies (Government of Saskatchewan, 2009). While a relatively small proportion of public water supplies (27%) use a surface water source; these primarily serve larger communities and are used by 57% of the population. Groundwater sources are used by 73% of public water supplies, which serve approximately 28% of the population (Government of Saskatchewan, 2009). Most private water

supplies are also based on groundwater, and it is estimated that there are over 66,000 private wells in use in Saskatchewan (Thompson, 2001).

Considering both public water supplies and private wells, it is estimated that 43% of the population obtains water from groundwater sources, primarily in rural and remote areas (Environment and Climate Change Canada, 2007). Previous investigations have identified concerns about groundwater quality in parts of Saskatchewan. A study of private and public wells reported that of 25 private wells and 36 wells operated by rural municipalities (RMs), 9 private wells and 5 RM wells exceeded the current drinking water standard for arsenic (10 µg/L) that is applied to regulated supplies (Thompson et al., 1999). In that study a maximum arsenic concentration of 117 µg/L was reported (Thompson et al., 1999). Of the 14 wells with arsenic concentrations exceeding the standard for regulated supplies, 9 were located in relatively close proximity to each other, raising concerns about the potential for “hot spots” of high arsenic concentrations in Saskatchewan groundwater (Thompson et al., 1999).

Other studies examined minerals in groundwater that may affect the aesthetic qualities of drinking water as well as cause damage to appliances and plumbing fixtures. High concentrations of iron, manganese, calcium, magnesium, and sulphate were common (Thompson, 2003).

Another study examined private wells for a wide range of parameters commonly monitored in regulated water supplies, and found that 99.6% of 535 wells exceeded at least one of Saskatchewan’s Drinking Water Quality Standards or Objectives (Sketchell and Shaheen, 2000). While most of these wells exceeded objectives that targeted the palatability of drinking water, 35% of wells exceeded a health-related standard (Sketchell and Shaheen, 2000).

### **1.3 Vulnerability of rural populations to water-related health**

Water supplies in rural and remote areas are vulnerable to problems with poor quality from both microbiological and chemical contamination. This is true for private water supplies (Corkal et al., 2004; Charrois, 2010) and small public supplies (Peterson and Torchia, 2008). Source water may be of poor quality due to high concentrations of naturally occurring substances such as heavy metals or minerals. In addition, rural supplies are also at risk from contamination from chemicals and microbes depending on surrounding agricultural or industrial land uses (Corkal et

al., 2004). Peterson and Torchia (2008) argue that the vulnerability of rural water supplies to contamination from run off means that these supplies require more intensive treatment than is typically available or in practice.

It is estimated that over 4 million Canadians rely on private water supplies, primarily private wells (Corkal et al., 2004). Researchers have highlighted common problems with groundwater sources in rural Saskatchewan (Thompson et al., 1999; Sketchell and Shaheen, 2000; Thompson, 2001; Thompson, 2003). However, these issues were not unique to Saskatchewan. For example, one study reported that 32% of private wells sampled on Alberta farms exceeded at least one health related standard and 93% exceeded at least one aesthetic guideline (Corkal et al., 2004). In the U.S., a survey of private wells across the U.S. found that 20% exceeded at least one health standard applied to regulated water supplies, and over 50% exceeded guidelines for substances that affect the aesthetic quality of the water (DeSimone et al., 2009). An examination of arsenic concentrations in wells in New Hampshire reported that private wells had significantly higher concentrations of arsenic than municipal wells (Peters et al., 1999).

Landowners with private water supplies bear the responsibility for ensuring the safety of their water supply, but this requires considerable knowledge of the risks as well as mitigation strategies (Corkal et al., 2004). Although data are scarce on the proportion of landowners with private wells that routinely test their water supplies, it appears that testing of private water supplies is infrequent and not comprehensive (Corkal et al., 2004; Charrois, 2010). A study in Ontario that surveyed users of private water supplies reported that over 20% of respondents had never tested their water (Jones et al., 2006). Testing for bacteria was most common (88% of respondents), while less than a quarter of respondents tested for other contaminants including heavy metals (Jones et al., 2006).

Rural residents who access regulated water supplies may still face water quality issues. Adequate water treatment requires systems designed to handle specific problems with source water quality as well as highly trained personnel, both of which can be logistically challenging in small communities and rural areas (Hrudey, 2008). It has also been suggested that private landowners and smaller public supplies are more likely to lack the resources to avoid placement of wells in

aquifers with poor water quality, including high concentrations of arsenic, than public systems serving larger populations (Focazio et al., 2000).

Water quality issues in rural areas are unique and multi-faceted. A better understanding of the potential risks posed by rural water supplies and effective strategies to mitigate these risks is critical to design policies to improve management of rural water supplies.

#### **1.4 Perceptions of Risk and Quality and Impacts on Consumption**

Despite the potential risks associated with water quality in rural areas, few studies have examined the perception of risks from drinking water among the rural population, the water consumption choices being made in rural areas, and how risk perception is related to consumption of household tap water. Understanding risk perception and behavior in response to that perception is critical in developing effective risk communication and policy (Slovic et al., 1982)

Water quality (e.g., taste, odor, color) and water safety (i.e. freedom from contaminants with an adverse effect on health) are often considered as separate constructs. However, when considering perceptions of quality and safety, these two constructs are inevitably intertwined. Understanding water quality and risk perception is essential for developing and communicating public health recommendations. It is also important to understand in the context of estimating exposures when personal protective measures can be taken in the face of perceived risk; for example, in-home treatment of drinking water to make it safer or to improve its palatability.

In general, risk perception is complex and depends on a number of factors including socioeconomic factors, education, gender, and social amplification through media or peers (Renn et al., 1992; Finucane et al., 2000; Dosman et al., 2001). While objective criteria may play a role in forming perceptions, ultimately risk perception is often based on intuitive judgments based on personal experiences and social influences (Slovic, 1987).

In the context of perceptions of water quality and risk, Anadu and Harding (2000) defined risk perception as “an individual’s subjective judgment (based on aesthetic and non-aesthetic qualities) about drinking water.” A review of perceptions of water quality and health risks from drinking water identified important factors that influence perceptions of drinking water,

including aesthetic qualities, prior experience, familiarity, interpersonal information, trust and control issues, and personal and cultural influences (Doria, 2010).

The aesthetic qualities of drinking water, particularly taste and odor, appear to be particularly important in perceptions of water quality and risk (Jardine et al., 1999; Jones et al., 2005; Jones et al., 2007; Doria et al., 2009). Arguably, the aesthetic characteristics of water are likely a poor indicator of risk, as many potentially harmful water contaminants would not be expected to change the aesthetic quality of water. By definition, Saskatchewan's drinking water quality objectives are meant to protect the palatability of drinking water for consumers, and are not considered health risks. Nevertheless, personal experience is thought to have a major influence on perceptions of water quality and risk, and taste and odor are the main metric by which consumers experience, and therefore make judgments, about their drinking water (Doria, 2010).

The paradox of taste versus safety was illustrated in a survey of residents of two communities in Quebec, where perception of water quality was compared to proxy measures of water quality including level of residual chlorine and distance from water treatment plant (Turgeon et al., 2004). Residents at the extremities of the distribution system generally perceived lower risks from their drinking water, and it was hypothesized that higher chlorine residuals nearer the treatment plant had a negative impact on taste ratings, lending support to the idea that poor taste ratings are associated with a higher perceived risk, regardless of the actual risk (Turgeon et al., 2004).

A sense of control over a hazard is thought to reduce the level of risk perceived (Doria 2010). Although residents with private water supplies bear sole responsibility for the safety and quality of their water supply, that responsibility could be accompanied by a sense of control that alleviates concern over the perceived risks from the water supply. For those residents accessing a public water supply, a lack of control may contribute to risk perception, and trust in the supplier becomes an important consideration (Doria et al., 2009; McSpirit and Reid 2011; Saylor et al., 2011). For consumers of both private and public water supplies, familiarity with the water supply likely plays an important role in risk perception (Dietrich, 2006; Doria, 2010).

Other factors, such as sex, age, and socioeconomic factors, also play a role in the perception of water quality and risk, although their relative contributions have been found to vary considerably



between communities and geographic regions (Turgeon et al., 2004; Doria 2010; McSpirit and Reid 2011). In Quebec, the effect of age was a significant risk factor in models of dissatisfaction with taste in two communities, but the direction was different in each community (Turgeon, 2004).

Studies examining factors that influence risk perception often produce contradictory findings, suggesting that factors predicting risk perception vary considerably depending on the study context (Dosman et al., 2001). Comparing studies of water quality and risk perception is difficult, primarily due to the use of different research instruments, but also due to geographic differences in perceptions (Doria et al., 2009). Furthermore, few studies have attempted to measure perceptions relating to water quality in rural areas.

According to Statistics Canada (Statistics Canada, 2011), about 20% of Saskatchewan residents reported using bottled water as their primary source of drinking water. The proportion of bottled water users in rural Saskatchewan is unknown. Perception of health risks from tap water have been found to be associated with the choice to drink bottled water (Doria et al., 2009; Dupont et al., 2010; Hu et al., 2011; Saylor et al., 2011), although aesthetic complaints about tap water also play a role (Levallois et al., 1999; Doria et al., 2009; Dupont et al., 2010; Saylor et al., 2011). Many other factors also influence the choice to drink bottled water including age, sex, income, household water source, and regional differences (Levallois et al., 1999; Doria et al., 2009; Dupont et al., 2010; Saylor et al., 2011). In rural and remote areas, the availability and cost of bottled water is an important influence the choice to drink bottled water (Doria, 2006).

Studies of factors influencing the choice of drinking water have largely been set in urban settings. Because populations in rural and remote areas might be more likely to use water supplies that are at risk of contamination and often not adequately tested, it is important to gain a better understanding how this population perceives risks from their drinking water and how these perceptions influence drinking water choices. A better understanding of perception of water quality and risk could help influence public health policy and education to promote safety of drinking water supplies. In addition, a better understanding of the choices rural residents make about consumption of their household tap water can inform future research examining links between exposure to contaminants in drinking water and health outcomes.

## **1.5 Associations between arsenic in drinking water and chronic disease**

### *1.5.1. Associations between arsenic and type 2 diabetes mellitus*

Evidence for an association between exposure to inorganic arsenic in drinking water and increased prevalence of type 2 diabetes has been examined in several recent reviews (Navas-Acien et al 2006; Chen et al., 2007; Maull et al., 2012; Kuo et al., 2013). These reviews compared studies in areas where high arsenic concentrations in drinking water were common to those in areas with low to moderate arsenic concentrations. However, areas characterized by high concentrations versus low to moderate exposure were defined by slightly different criteria. High arsenic exposure areas were defined as areas where arsenic concentrations in groundwater were typically above 100 µg/L in one review (Navas-Acien et al., 2006) while the U. S. National Toxicology Program workshop review (Maull et al., 2012) defined high arsenic exposure areas to be those with groundwater arsenic concentrations greater than 150 µg/L. Areas with typical arsenic concentrations below these thresholds were considered areas of low to moderate exposures. The review by Kuo et al. (2013) updated the review by Maull et al. (2012) and used the 150 µg/L threshold to define high exposures although Chen et al. (2007) did not explicitly define a threshold.

Studies included in several reviews (Navas-Acien et al., 2006; Chen et al., 2007; Maull et al., 2012) were primarily done in high arsenic exposure areas. These reviews consistently concluded that where arsenic concentrations in drinking water was high, the evidence supported an association between increased exposure to inorganic arsenic in drinking water and increased risk of diabetes. Maull et al. (2012) cautioned that despite the consistency of results in studies where arsenic concentrations were typically >150 µg/L, the evidence was somewhat limited by the cross-sectional nature of the studies cited, uncertainty in outcome measures, and a lack of individual measures of exposure. Navas-Acien et al. (2006), Chen et al. (2007) and Maull et al. (2012) concluded that at low to moderate arsenic exposures the evidence was not sufficient to reach a conclusion about an association between arsenic and diabetes.

A review by Kuo et al. (2013) updated the National Toxicology Program workshop review (Maull et al., 2012) and examined evidence from recent studies including those done in areas characterized by low to moderate arsenic exposures. This review cited two prospective studies

performed in U.S. populations (James et al., 2013; Kim et al., 2013) as supporting a temporal association between exposure to arsenic and development of diabetes. While one of these studies reported a statistically significant association between arsenic exposure and diabetes (James et al., 2013), the other was only suggestive of such an association (Kim et al., 2013), and both studies were relatively small.

James et al. (2013) used an exposure matrix approach to estimate a time weighted arsenic exposure metric based on groundwater arsenic concentrations (measured or predicted by geostatistical methods). Lifetime residential, workplace and school locations for the exposure matrix were reconstructed by interviews or county records and drinking water consumption at each location by interview, leaving potential for recall bias in the exposure estimation. Whereas, Kim et al. (2013) relied on urinary arsenic measured at the beginning of the study as an exposure estimate for a case-control study of participants who developed diabetes during the study period compared to controls, which could have resulted in misclassification of long-term exposures.

A cross sectional study in the U.S. reported significant associations between urinary arsenic concentrations and diabetes prevalence (Gribble et al., 2012), but other recent studies reported that exposure to low to moderate concentrations of arsenic in drinking water was not associated with differences in the prevalence of diabetes (Chen et al., 2010; Makris et al., 2012; Li et al., 2013). Chen et al. (2010) evaluated time weighted arsenic exposures as well as urinary arsenic exposures and adjusted for dietary sources. Makris et al. (2012) evaluated cumulative arsenic exposures based on exposures from wells with a maximum arsenic concentration of 70 µg/L, while Li et al. (2013) compared exposures to wells with arsenic concentrations < 10 µg/L to those with concentrations of 10-50 µg/L and > 50 µg/L.

A meta-analysis of 17 published observational studies of associations between inorganic arsenic and diabetes has also been reported (Wang et al., 2014). Studies from both high and moderate to low arsenic areas were included in the meta-analysis. Separate meta-analyses were performed for studies where arsenic exposure was estimated from drinking water arsenic concentrations and studies where arsenic exposure was estimated by urinary concentrations of arsenic. The pooled analysis of results in both groups of studies demonstrated statistically significant associations between increased exposure to arsenic and diabetes. Four studies were used in a dose-response

analysis, which demonstrated that a 100 µg/L increase in arsenic concentration in drinking water was associated with a 13% increase in diabetes prevalence (Wang et al., 2014).

The necessity of separating observational studies based on drinking water arsenic concentrations from those based on urinary arsenic concentrations for the meta-analysis by Wang et al. (2014) illustrates a considerable challenge in the comparison of studies of associations between arsenic exposure and diabetes. The variety of techniques used to estimate exposure to arsenic in various studies include detailed individual-level cumulative exposure estimates (e.g., Chen et al., 2010; James et al., 2013), arsenic concentrations in drinking water (e.g., Meliker et al., 2007; Del Razo et al., 2011), residence in an high arsenic exposure or arsenicosis-endemic area (e.g., Tsai et al., 1999; Wang et al., 2003), presence of keratosis as an indicator of arsenic exposure (e.g., Rahman et al., 1998), and biomarkers for arsenic exposure (hair, serum, or urine arsenic concentrations) (e.g., Afridi et al., 2008; Serdar et al., 2009; Gribble et al., 2012; Kim et al., 2013).

Accurate estimation of arsenic exposure is problematic, especially over long time periods. Individual cumulative exposure estimates may rely on participant recall of residence locations and drinking water intake combined with estimates of arsenic concentrations in groundwater in those locations (e.g., James et al., 2013) and misclassification bias is a concern. Biomarkers estimate relatively recent individual exposures, and unless speciation of arsenic metabolites is done, it can be difficult to differentiate routes of exposure of arsenic (e.g., water, dietary, occupational). Difficulties in accurately estimating arsenic exposure might contribute to inconsistent results in studies of the effects of low to moderate concentrations. The inability to accurately discriminate between relative small differences in exposure levels could mask small differences in diabetes prevalence associated with exposure.

The challenge of estimating accurate arsenic exposures is exacerbated by a poor understanding of the relevant exposure or induction period for arsenic on diabetes. While several mechanisms have been proposed for the influence of arsenic on the development of diabetes (Tseng, 2004), the clinical importance of the mechanisms has not been fully elucidated. A poor understanding of the salient induction period for arsenic complicates the estimation of arsenic exposures, and likely represents a form of non-differential misclassification that biases results of studies of associations between arsenic and chronic disease towards the null (Rothman, 1981).

Comparison of studies is also hindered by heterogeneous methodology with respect to case definition, including self reported, fasting blood glucose measurements, glucose tolerance testing, or medication/treatment history. In addition, control of important confounders is not uniform among studies. While many studies control for age and sex, the range of other important factors such as body mass index, lifestyle habits, and socioeconomic status considered as confounders varied considerably between studies.

Ultimately, challenges in comparing previous studies and their conflicting results makes it difficult for public health professionals and policy makers to assess risks from low to moderate exposures to arsenic on diabetes prevalence. With the incidence and prevalence of diabetes increasing in Saskatchewan, especially among First Nations populations (Dyck et al., 2010), identifying potential modifiable risk factors for diabetes is important for reducing the burden of disease in Saskatchewan.

#### *1.5.2 Associations between arsenic and cardiovascular disease*

Reviews of studies examining associations between arsenic exposure and hypertension and cardiovascular disease have concluded that evidence from observational studies supports associations between high concentrations of arsenic in drinking water and hypertension (Abir et al., 2011; Abhyankar et al., 2012), as well as ischemic heart disease (IHD) and stroke (Navas-Acien et al., 2005; Wang et al., 2007; Moon et al., 2012; Tsuji et al., 2014). As with type 2 diabetes, the evidence is less compelling for associations between exposures to low to moderate concentrations of arsenic in drinking water and cardiovascular disease.

A systematic review of studies examining associations between drinking water arsenic and hypertension (Abhyankar et al., 2012) defined high drinking water arsenic concentrations as those above 50 µg/L, slightly lower than the threshold used by reviews of studies examining associations between arsenic and diabetes. Results were inconsistent, but studies characterized by both high and low arsenic exposure reported associations between arsenic exposure and hypertension. Inconsistency in the reported results could be related to heterogeneity in study methodology. The absence of prospective studies was cited as a primary reason for the inability to reach conclusions regarding a causal association between arsenic and hypertension (Abhyankar et al., 2012).

A meta-analysis of studies examining associations between arsenic and hypertension was also published (Abir et al., 2012). The results of this analysis supported an association between exposure to arsenic in drinking water and hypertension. However, the authors cautioned that the small number of studies, which were primarily cross sectional and of variable quality, limited the ability to make any causal inferences regarding arsenic and hypertension (Abir et al., 2012).

Recent studies in the U.S. have produced conflicting results. A study in Texas demonstrated an association between arsenic concentrations in groundwater and hypertension in an area where the estimated median groundwater concentration was 6.5 µg/L (Gong and O'Bryant, 2012). Other studies were based on biomarker estimates of exposure. Toenail arsenic concentrations were associated with an increase in systolic blood pressure (Mordukhovich et al., 2012), but in another study urinary arsenic concentrations were not associated with hypertension (Jones et al., 2011).

While relatively similar endpoints were measured in studies examining associations between arsenic and hypertension, there was some heterogeneity in the methods used to define hypertension (Abir et al., 2011; Abhyankar et al., 2012). Case definitions were often based on blood pressure measurements, but among these, the systolic and diastolic values used to define hypertension varied. Other studies depended on hypertension cases being self-reported or were based on the use of antihypertensive medication.

In contrast, the endpoints measured in studies examining associations between arsenic exposure and cardiovascular disease were diverse (Navas-Acien et al., 2005; Wang et al., 2007; Moon et al., 2012; Tsuji et al., 2014). The term “cardiovascular disease” is a more heterogeneous term that includes a variety of outcomes including ischemic heart disease, cerebrovascular disease, and peripheral vascular disease. This heterogeneity was reflected in the variety of endpoints measured in studies examining associations between arsenic exposure and cardiovascular disease. Studies have used endpoints defined by subclinical indicators of CVD such as electrocardiogram changes (e.g., Mordukhovich et al., 2009), carotid intima-media thickness (e.g., Chen et al., 2006), and carotid atherosclerosis (e.g., Wu et al., 2006). Other studies have used a range of clinical presentations of CVD including ischemic heart disease and stroke, although some focussed on mortality and others on morbidity. Within this range of endpoints, methods for defining cases also vary; for example, the occurrence of CVD might be self-reported

(e.g., Gong and O'Bryant, 2012) or defined from death certificates (e.g., Meliker et al., 2007), hospitalization records (Lisabeth et al., 2010), or from speaking with family members of a deceased patient (e.g., Chen et al., 2011).

A review of studies examining associations between low-level arsenic exposure and cardiovascular diseases, intended to inform a U.S. Environmental Protection Agency non-cancer reference dose for arsenic for toxicological risk assessment, reported that epidemiological studies supported a no-adverse-effect-level arsenic concentration of 100 µg/L in drinking water (Tsuji et al., 2014).

Among U.S. studies, conflicting results have been reported, but several recent studies have found evidence suggestive of associations of low to moderate concentrations of arsenic in drinking water and CVD.

A study using toenail arsenic concentrations as a biomarker for arsenic exposure in an area with very low drinking water arsenic concentrations reported a significant association between toenail arsenic concentrations and electrocardiograph abnormalities (Mordukhovich et al., 2009). Lisabeth et al (2010) found an association between drinking water arsenic concentrations and hospital admissions for ischemic stroke in an area with a median drinking water arsenic concentration of 1.83 µg/L. A study in an area with a median drinking water arsenic concentration of 6.5 µg/L found an association between arsenic concentration and increased risk of coronary heart disease (Gong and O'Bryant, 2012). Moon et al. (2013) reported significant associations between arsenic exposure measured by urinary arsenic concentrations and the incidence of cardiovascular disease, coronary heart disease, and stroke in areas where the maximum groundwater arsenic concentration was 61 µg/L, although the effect estimates were attenuated when adjusted for confounders. In a case-cohort study in an area with low to moderate (10-100 µg/L) groundwater arsenic concentrations, time weighted arsenic exposure (estimated using an exposure matrix approach) was associated with incidence of coronary heart disease (James et al., 2015).

While some U.S. studies have produced evidence suggestive of a link between low to moderate arsenic exposure and cardiovascular disease, caution is warranted due to the cross-sectional nature of most studies. Several cohort studies from areas with high groundwater arsenic

concentrations demonstrated associations between arsenic and CVD only for arsenic concentrations > 50 µg/L as summarized in the review by Tsuji et al (2014).

The considerable uncertainty associated with inconsistent results for measures of association between low exposures to arsenic and CVD is exacerbated by the heterogeneity of outcomes investigated as well as in the methods used to estimate arsenic exposures. Assessing exposure is also made difficult by the potential for a long latent period between exposure and outcome. It has been suggested that epigenetic effects of arsenic are important in the development of CVD, and that *in utero* exposure to high arsenic concentrations can predispose to CVD later in life (Smith and Steinmaus, 2009; Farzan et al., 2013; Abdul et al., 2015). While many physiological effects of arsenic on the cardiovascular system have been reported in experimental studies (Tsuji et al., 2014; Abdul et al., 2015), a better understanding how arsenic mediates the development of CVD is needed (Moon et al., 2012), particularly for establishing the appropriate induction period over which exposure should be estimated.

Failure to control for confounding variables, or consider if included risk factors are true confounders or mediators, is a potential source of bias and is an important contributor to heterogeneity in the quality of studies and effect estimates of associations between exposures and outcomes (O'Connor and Sargeant, 2014). Comparison of studies that have investigated associations between arsenic and cardiovascular disease is also made more difficult by the variety of variables considered as confounders in these studies (Navas-Acien et al., 2005; Wang et al., 2007; Moon et al., 2012; Tsuji et al., 2014).

As with associations between arsenic and diabetes, uncertainty about the risks of arsenic exposures at the concentrations that have been reported for groundwater in Saskatchewan is a challenge for public health professionals and policy-makers. Given the very high burden of disease contributed by cardiovascular diseases in Canada (Public Health Agency of Canada, 2014), further research on potential risk factors that could be modified to reduce the burden of CVD is indicated.

### *1.5.3 Effect modifiers on associations between arsenic and chronic disease*

Recent research also suggests that genetic susceptibility, including polymorphisms in the AS3MT gene (Hsieh et al., 2011; Gong and O'Bryant, 2012; Gonzalez-Horta et al., 2012;



Drobná et al., 2013) is important in mediating the effects of arsenic on the development of diabetes and CVD. Therefore, certain populations may be more likely to experience adverse effects of arsenic in drinking water. In addition, it appears malnutrition, particularly folate deficiency, increase the susceptibility of individuals to the adverse effects of arsenic (Tsuji et al., 2014), which may influence the importance of arsenic as a risk factor for diabetes and CVD in different populations.

Given the uncertainty about the adverse effects of low to moderate concentrations of arsenic in drinking water, coupled with the possibility of interactions between inorganic arsenic and genetic and dietary factors, it is extremely difficult to extrapolate the results of other studies to the population of rural Saskatchewan. While concerns have been raised about the presence of arsenic above recommended drinking water standards in some parts of Saskatchewan, no studies have been done to assess the relationship of arsenic in Saskatchewan groundwater and the prevalence or incidence of chronic disease.

### **1.6 Indirect effects of Poor Quality Drinking Water on Diabetes and Cardiovascular Disease**

In addition to the direct effects of drinking water contaminants on the pathogenesis of diabetes and CVD, poor quality drinking water could indirectly lead to decreased consumption of water and increased consumption of sugar sweetened beverages. The influence of perceptions of water quality and risk on the relative consumption of water and sugar sweetened beverages has not been studied in detail. A recent U.S. study found that the perception that tap water was unsafe was associated with increased consumption of sugar-sweetened beverages among some minority groups (Onufrak et al., 2014). It also possible that where people find the household water unpalatable, consumption of other beverages including sugar-sweetened beverages could be increased.

Adequate consumption of water has been identified as important in management of health body weight (Stookey et al., 2008; Daniels and Popkin, 2010). In addition, consumption of sugar-sweetened beverages has been linked to increase rates of obesity and overweight (Schulze, 2004; Malik et al., 2006; Hu and Malik, 2010), which are important risk factors for the development of type 2 diabetes, hypertension, and CVD among other chronic diseases (Kopelman, 2000; Field et

al., 2001). Furthermore, increased consumption of sugar-sweetened beverages was associated with increased rates of diabetes and cardiovascular disease, independent of effects mediated by increased rates of obesity (Hu and Malik, 2010; Malik et al., 2010a, 2010b).

Given previous studies that have identified high concentrations of contaminants affecting palatability in Saskatchewan groundwater (Sketchell and Shaheen, 2000; Thompson, 2003), residents of rural and remote areas of Saskatchewan may be at greater risk of having household tap water with poor palatability compared to those residing in urban centers. While the literature is not clear on the effects of poor water palatability and consumption of water or sugar-sweetened beverages, residents of rural areas may be more vulnerable to any indirect impacts poor water palatability on the rates of chronic disease

### **1.7. Assessing effects of mixtures of contaminants in drinking water**

There is increasing recognition that environmental contamination consists of complex mixtures that act in concert, and that studying the health effects of mixtures in drinking water may be a more sensible approach than studying effects of single contaminants (Monosson, 2005; Schwarzenbach et al., 2006; Murphy et al., 2012). While laboratory methods are being developed to assess complex mixtures (Murphy et al., 2012), multivariate techniques represent a way to screen groups of parameters summarized from currently available water quality data for health effects (Burstyn, 2004; Villanueva et al., 2012). Principal components analysis has been applied to water quality monitoring data to characterize water quality over large regions using geostatistical techniques (Sánchez-Martos et al., 2001; Satyaji Rao et al., 2009; Shyu et al., 2011; Nazzal et al., 2015). However, no studies were identified that combined the use of multivariate techniques with geostatistics to estimate exposures for use in epidemiological studies investigating associations between mixtures of drinking water contaminants and health outcomes.

### **1.8 Objectives of research**

While Saskatchewan's rural residents are vulnerable to water-related health risks, there have been no previous studies examining the perception of these risks among this population and how risk perception affects water consumption patterns. This research aimed to establish a baseline understanding of key issues related to drinking water in Saskatchewan's rural communities,

especially sources of drinking water, drinking water choices, and perceptions of health risks from drinking water. While water quality issues are especially important for Indigenous communities (Ekos Research Associates, 2011), this research focussed more generally on rural communities because a concurrent Saskatchewan study targeted specific concerns about water quality in Indigenous communities.

In addition, studies in other populations have identified associations between drinking water with elevated levels of arsenic and type 2 diabetes and cardiovascular disease. In Saskatchewan, particularly in rural areas where groundwater is a common drinking water source, naturally occurring water contaminants including arsenic have the potential to contribute to poor health outcomes. However, no studies examining potential associations between water quality and chronic disease in Saskatchewan were identified. In addition, no epidemiological studies addressing associations between mixtures of contaminants and health outcomes were identified.

The overall hypothesis of this thesis is that poor water quality in some rural communities in Saskatchewan could be associated with an increased risk of important chronic diseases. The objectives addressed in this thesis address important gaps in knowledge about perception of risks from drinking water among rural residents, drinking water preferences in rural populations, and the potential impacts of water quality on the occurrence of chronic disease in Saskatchewan.

#### *1.8.1 Investigate factors influencing perceptions of water quality and safety among rural Saskatchewan residents*

The first research chapter examined the level of concern with drinking tap water in rural Saskatchewan. People value safe drinking water, but risk perception related to drinking water is complex. The hypothesis for the first research chapter in the thesis was that numerous factors influence perceptions of water quality and health risks from drinking water among rural Saskatchewan residents, and that these factors would vary with the type of water source used.

The objective for this chapter was to describe the occurrence of, and to evaluate risk factors associated with, having an aesthetic complaint about the tap water, the belief that the tap water was safe to drink, the belief that the tap water had made someone ill, and fear that the household's water source would become contaminated.

A better understanding of perceptions of household water quality and safety can inform public health messaging about water safety and management in rural areas.

#### *1.8.2 Factors affecting choices to drink tap water or bottled water and to treat household tap water*

The second research chapter explored factors influencing exposure to tap water. Perceptions of water quality and safety are likely to influence choices for water consumption, and some of the factors that influence drinking water choices are likely unique among rural populations.

It was hypothesized that the type of water source used, as well as perceptions of quality and risk, would influence choices about water consumption. The objective was to evaluate the risk factors associated with the choice to consume tap water on a regular basis, the choice to consume primarily bottled water, and the choice to use in-home water treatment equipment in rural Saskatchewan.

A better understanding of how water sources and water quality and risk perception might influence choices around drinking water in rural Saskatchewan can inform public health messaging as well as indicate what measures people take to modify their exposures to perceived risks from drinking water.

#### *1.8.3 Summarize existing surveillance data about drinking water quality*

The third research chapter focussed on describing the quality of the water available to residents of rural Saskatchewan. Water quality surveillance data from public water systems monitored by the Water Security Agency was identified as a source of information on the exposure of residents who access public water supplies supplied by groundwater sources within the province. Because groundwater-based systems are almost exclusively used outside of urban environments, information about these systems represent the water quality experienced by residents of smaller communities that rely on wells for drinking water in rural areas. Similarly, data collected as part of the Rural Water Quality Advisory Program provides a unique and rich source of information about water quality in private wells in rural Saskatchewan.

Because many different parameters are monitored as health standards and aesthetic objectives, principal components analysis was identified as a potential tool to summarize information from

groups of contaminants. In addition, there was a need to identify the best method of predicting arsenic concentrations and principal component scores for geographic regions of interest by comparing the performance of ordinary kriging, universal kriging, and empirical Bayesian kriging.

The overall objective was to evaluate the potential use of multivariate statistics combined with geostatistics to summarize existing groundwater monitoring data from public water supplies and private wells. Maps summarizing arsenic concentrations and patterns of poor water quality and palatability could be useful resources in prioritizing parameters for regulatory testing of small public water supplies and in public education campaigns to encourage private water supply users to test water appropriately. In addition, data summarized using these techniques can be used for exposure assessment in studies investigating health impacts of water quality in rural areas.

#### *1.8.4 Investigate associations between water quality and type 2 diabetes mellitus*

The fourth chapter examined the potential for an association between local water quality and the occurrence of type 2 diabetes mellitus. Three separate objectives were developed to explore this question:

- i. Investigate associations between arsenic concentrations in drinking water from groundwater sources and type 2 diabetes.
- ii. Investigate associations between principal component scores for health standards, reflecting poor water quality, and type 2 diabetes.
- iii. Investigate associations between principal component scores for aesthetic objectives, reflecting water with poor palatability, and type 2 diabetes.

The exposure assessment used for investigating these associations was informed by the summary of existing water surveillance data outlined in objective 1.8.3.

The first two objectives consider the potential for direct effects on the incidence and prevalence of diabetes by arsenic and contaminants identified as health-based drinking water standards. The third objective evaluates aesthetic objective parameters and the indirect effects of water with poor palatability on the incidence and prevalence of diabetes.

#### *1.8.5 Investigate associations between water quality and cardiovascular disease*

The fifth research chapter investigates the potential association between local water quality and the risks for cardiovascular disease. Similar to objective 1.8.4, three objectives made up this component of the thesis:

- i. Investigate associations between arsenic concentrations in drinking water from groundwater sources and hypertension, ischemic heart disease, and stroke.
- ii. Investigate associations between principal component scores for health standards, reflecting poor water quality, and hypertension, ischemic heart disease, and stroke.
- iii. Investigate associations between principal component scores for aesthetic objectives, reflecting water with poor palatability, and hypertension, ischemic heart disease, and stroke.

Associations between water quality, as summarized in objective 1.8.3, and prevalence of hypertension, and ischemic heart disease and stroke were investigated. Both the direct effects of arsenic and contaminants identified as health standards as well as the indirect effects of aesthetic objectives on the prevalence of cardiovascular disease were evaluated.

Direct and indirect impacts of water quality on diabetes and cardiovascular disease could represent modifiable risk factors for which interventions could be implemented, contributing to the mitigation of the health and economic burden of these diseases.

## 1.9 References

- Abdul, K.S.M., Jayasinghe, S.S., Chandana, E.P.S., Jayasumana, C., De Silva, P.M.C.S., 2015. Arsenic and human health effects: A review. *Environ. Toxicol. Pharmacol.* 40, 828–846. doi:10.1016/j.etap.2015.09.016
- Abhyankar, L.N., Jones, M.R., Guallar, E., Navas-Acien, A., 2012. Arsenic Exposure and Hypertension: A Systematic Review. *Environ. Health Perspect.* 120, 494–500.
- Abir, T., Rahman, B., D’Este, C., A, F., Milton, A.H., 2012. The Association between Chronic Arsenic Exposure and Hypertension: A Meta-Analysis. *J. Toxicol.* 2012, 1–13. doi:10.1155/2012/198793
- Afridi, H.I., Kazi, T.G., Kazi, N., Jamali, M.K., Arain, M.B., Jalbani, N., Baig, J.A., Sarfraz, R.A., 2008. Evaluation of status of toxic metals in biological samples of diabetes mellitus patients. *Diabetes Research and Clinical Practice* 80, 280–288. doi:10.1016/j.diabres.2007.12.021
- Anadu, E.C., Harding, A.K., 2000. Risk perception and bottled water use. *J. Am. Water Works Assoc.* 92, 82–92.
- Burstyn, I., 2004. Principal Component Analysis is a Powerful Instrument in Occupational Hygiene Inquiries. *Ann. Occup. Hyg.* 48, 655–661. doi:10.1093/annhyg/meh075
- Charrois, J.W.A., 2010. Private Drinking Water Supplies: Challenges for Public Health. *Can. Med. Assoc. J.* 182, 1061–1064. doi:10.1503/cmaj.090956
- Chen, Y., Hakim, M.E., Parvez, F., Islam, T., Rahman, A.M., Ahsan, H., 2006. Arsenic exposure from drinking-water and carotid artery intima-medial thickness in healthy young adults in Bangladesh. *Journal of Health, Population and Nutrition* 24, 253–257.
- Chen, C.-J., Wang, S.-L., Chiou, J.-M., Tseng, C.-H., Chiou, H.-Y., Hsueh, Y.-M., Chen, S.-Y., Wu, M.-M., Lai, M.-S., 2007. Arsenic and diabetes and hypertension in human populations: A review. *Toxicol. Appl. Pharmacol.* 222, 298–304. doi:10.1016/j.taap.2006.12.032
- Chen, Y., Ahsan, H., Slavkovich, V., Peltier, G.L., Gluskin, R.T., Parvez, F., Liu, X., Graziano, J.H., 2010. No association between arsenic exposure from drinking water and diabetes mellitus: a cross-sectional study in Bangladesh. *Environ. Health Perspect.* 118, 1299–1305.
- Chen, Y., Graziano, J.H., Parvez, F., Liu, M., Slavkovich, V., Kalra, T., Argos, M., Islam, T., Ahmed, A., Rakibuz-Zaman, M., others, 2011. Arsenic exposure from drinking water and mortality from cardiovascular disease in Bangladesh: prospective cohort study. *Bmj* 342, d2431.
- Corkal, D., Schutzman, W.C., Hilliard, C.R., 2004. Rural water safety from the source to the on-farm tap. *J. Toxicol. Environ. Health A* 67, 1619–1642.
- Daniels, M.C., Popkin, B.M., 2010. Impact of water intake on energy intake and weight status: a systematic review. *Nutr. Rev.* 68, 505–521. doi:10.1111/j.1753-4887.2010.00311.x
- Del Razo, L.M., García-Vargas, G.G., Valenzuela, O.L., Castellanos, E.H., Sánchez-Peña, L.C., Currier, J.M., Drobná, Z., Loomis, D., Stýblo, M., others, 2011. Exposure to arsenic in

- drinking water is associated with increased prevalence of diabetes: a cross-sectional study in the Zimapán and Lagunera regions in Mexico. *Environ. Health* 10, 73–83.
- DeSimone, L.A., Hamilton, P.A., Gilliom, R.J., 2009. Quality of ground water from private domestic wells. *Water Well J* 1, 33–37.
- Dietrich, A.M., 2006. Aesthetic issues for drinking water. *J. Water Health* 4, 11–16.
- Doria, M. de F., 2010. Factors influencing public perception of drinking water quality. *Water Policy* 12, 1–19. doi:10.2166/wp.2009.051
- Doria, M. de F., Pidgeon, N., Hunter, P.R., 2009. Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study. *Sci. Total Environ.* 407, 5455–5464.
- Doria, M.de F., 2006. Bottled water versus tap water: understanding consumers' preferences. *J. Water Health* 4, 271–276.
- Dosman, D.M., Adamowicz, W.L., Hrudey, S.E., 2001. Socioeconomic determinants of health- and food safety-related risk perceptions. *Risk Anal.* 21, 307–318. doi:10.1111/0272-4332.212113
- Drobná, Z., Del Razo, L.M., García-Vargas, G.G., Sánchez-Peña, L.C., Barrera-Hernández, A., Stýblo, M., Loomis, D., 2013. Environmental exposure to arsenic, AS3MT polymorphism and prevalence of diabetes in Mexico. *J. Expo. Sci. Environ. Epidemiol.* 23, 151–155.
- Dupont, D., Adamowicz, W.L., Krupnick, A., 2010. Differences in water consumption choices in Canada: the role of socio-demographics, experiences, and perceptions of health risks. *J. Water Health* 8, 671–686.
- Dyck, R., Osgood, N., Lin, T.H., Gao, A., Stang, M.R., 2010. Epidemiology of diabetes mellitus among First Nations and non-First Nations adults. *Can. Med. Assoc. J.* 182, 249–256. doi:10.1503/cmaj.090846
- Ekos Research Associates, 2011. Perceptions of Drinking Water Quality in First Nations Communities and General Population. Ottawa, ON.
- Environment and Climate Change Canada, 2007. Environment and Climate Change Canada - Water - Groundwater. URL <https://www.ec.gc.ca/eau-water/default.asp?lang=En&n=300688DC-1#sub5>.
- Farzan, S.F., Karagas, M.R., Chen, Y., 2013. In utero and early life arsenic exposure in relation to long-term health and disease. *Toxicol. Appl. Pharmacol.* 272, 384–390. doi:10.1016/j.taap.2013.06.030
- Field, A.E., Coakley, E.H., Must, A., Spadano, J.L., Laird, N., Dietz, W.H., Rimm, E., Colditz, G.A., 2001. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Arch. Intern. Med.* 161, 1581–1586.
- Finucane, M.L., Slovic, P., Mertz, C.K., Flynn, J., Satterfield, T.A., 2000. Gender, race, and perceived risk: the “white male” effect. *Health Risk Soc.* 2, 159–172. doi:10.1080/713670162
- Focazio, M.J., Welch, A.H., Watkins, S.A., Helsel, D.R., Horn, M.A., 2000. A retrospective analysis on the occurrence of arsenic in ground-water resources of the United States and



- limitations in drinking-water-supply characterizations (No. 99–4279), Water-resources investigations Report.
- Gong, G., O'Bryant, S.E., 2012. Low-level arsenic exposure, AS3MT gene polymorphism and cardiovascular diseases in rural Texas counties. *Environ. Res.* 113, 52–57. doi:10.1016/j.envres.2012.01.003
- Gonzalez-Horta, C., Sánchez-Ramírez, B., Ballinas-Casarrubias, L., Ishida-Gutiérrez, C., Del Razo, L.M., Garcia-Vargas, G., Drobná, Z., Loomis, D., Styblo, M., 2012. S-052: Exposure to Arsenic, AS3MT Genotype and Prevalence of Diabetes: Recent Evidence from Studies in Mexico. *Epidemiology* 23, S-306.
- Government of Saskatchewan, 2009. Sask H2O - Water Information Quick Facts. URL [http://www.saskh2o.ca/WaterInformation\\_QuickFacts.asp](http://www.saskh2o.ca/WaterInformation_QuickFacts.asp).
- Government of Saskatchewan, 1966. The Ground Water Regulations. URL <http://www.qp.gov.sk.ca/documents/English/Regulations/Regulations/SR172-66.pdf>
- Government of Saskatchewan, 2002a. The Water Regulations, 2002. URL [http://www.saskh2o.ca/DWBinder/Water\\_Regs\\_e10-21r1.pdf](http://www.saskh2o.ca/DWBinder/Water_Regs_e10-21r1.pdf)
- Government of Saskatchewan, 2002b. The Health Hazard Regulations. URL <http://www.qp.gov.sk.ca/documents/english/Regulations/Regulations/p37-1r10.pdf>
- Gribble, M.O., Howard, B.V., Umans, J.G., Shara, N.M., Francesconi, K.A., Goessler, W., Crainiceanu, C.M., Silbergeld, E.K., Guallar, E., Navas-Acien, A., 2012. Arsenic Exposure, Diabetes Prevalence, and Diabetes Control in the Strong Heart Study. *Am. J. Epidemiol.* 176, 865–874. doi:10.1093/aje/kws153
- Health Canada, 2014. Guidelines for Canadian Drinking Water Quality - Summary Table. URL [http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/sum\\_guide-res\\_recom/index-eng.php](http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/sum_guide-res_recom/index-eng.php).
- Hrudey, S.E., 2008. Safe water? Depends on where you live! *Can. Med. Assoc. J.* 178, 975–975. doi:10.1503/cmaj.080374
- Hsieh, Y.-C., Lien, L.-M., Chung, W.-T., Hsieh, F.-I., Hsieh, P.-F., Wu, M.-M., Tseng, H.-P., Chiou, H.-Y., Chen, C.-J., 2011. Significantly increased risk of carotid atherosclerosis with arsenic exposure and polymorphisms in arsenic metabolism genes. *Environ. Res.* 111, 804–810. doi:10.1016/j.envres.2011.05.003
- Hu, F.B., Malik, V.S., 2010. Sugar-sweetened beverages and risk of obesity and type 2 diabetes: Epidemiologic evidence. *Physiol. Behav.* 100, 47–54. doi:10.1016/j.physbeh.2010.01.036
- Hu, Z., Morton, L.W., Mahler, R.L., 2011. Bottled water: United States consumers and their perceptions of water quality. *Int. J. Environ. Res. Public Health* 8, 565–578.
- James, K.A., Byers, T., Hokanson, J.E., Meliker, J.R., Zerbe, G.O., Marshall, J.A., 2015. Association between Lifetime Exposure to Inorganic Arsenic in Drinking Water and Coronary Heart Disease in Colorado Residents. *Environ. Health Perspect.* 123, 128–134. doi:10.1289/ehp.1307839
- James, K.A., Marshall, J.A., Hokanson, J.E., Meliker, J.R., Zerbe, G.O., Byers, T.E., 2013. A case-cohort study examining lifetime exposure to inorganic arsenic in drinking water and diabetes mellitus. *Environ. Res.* 123, 33–38. doi:10.1016/j.envres.2013.02.005

- Jardine, C.G., Gibson, N., Hrudey, S.E., 1999. Detection of odour and health risk perception of drinking water. *Water Sci. Technol.* 40, 91–98.
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Mathews, E., Carr, D.J., Henson, S.J., 2006. Public perceptions of drinking water: a postal survey of residents with private water supplies. *BMC Public Health* 6, 94–104. doi:10.1186/1471-2458-6-94
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Henson, S.J., Mathews, E., 2007. A qualitative exploration of the public perception of municipal drinking water. *Water Policy* 9, 425–438. doi:10.2166/wp.2007.019
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Henson, S.J., Mathews, E., 2005. Public perception of drinking water from private water supplies: focus group analyses. *BMC Public Health* 5, 129. doi:10.1186/1471-2458-5-129
- Jones, M.R., Tellez-Plaza, M., Sharrett, A.R., Guallar, E., Navas-Acien, A., 2011. Urine arsenic and hypertension in US adults: the 2003-2008 National Health and Nutrition Examination Survey. *Epidemiol. Camb. Mass* 22, 153–161. doi:10.1097/EDE.0b013e318207fdf2
- Kim, N.H., Mason, C.C., Nelson, R.G., Afton, S.E., Essader, A.S., Medlin, J.E., Levine, K.E., Hoppin, J.A., Lin, C., Knowler, W.C., others, 2013. Arsenic exposure and incidence of type 2 diabetes in Southwestern American Indians. *Am. J. Epidemiol.* 177, 962–969.
- Kopelman, P.G., 2000. Obesity as a medical problem. *Nature* 404, 635–643.
- Kuo, C.-C., Moon, K., Thayer, K.A., Navas-Acien, A., 2013. Environmental chemicals and type 2 diabetes: an updated systematic review of the epidemiologic evidence. *Curr. Diab. Rep.* 13, 831–849.
- Levallois, P., Grondin, J., Gingras, S., 1999. Evaluation of consumer attitudes on taste and tap water alternatives in Quebec. *Water Sci. Technol.* 40, 135–139.
- Li, X., Li, B., Xi, S., Zheng, Q., Lv, X., Sun, G., 2013. Prolonged environmental exposure of arsenic through drinking water on the risk of hypertension and type 2 diabetes. *Environ. Sci. Pollut. Res. Int.* 20, 8151–8161. doi:10.1007/s11356-013-1768-9
- Lisabeth, L.D., Ahn, H.J., Chen, J.J., Sealy-Jefferson, S., Burke, J.F., Meliker, J.R., 2010. Arsenic in drinking water and stroke hospitalizations in Michigan. *Stroke* 41, 2499–2504.
- Makris, K.C., Christophi, C.A., Paisi, M., Ettinger, A.S., 2012. A preliminary assessment of low level arsenic exposure and diabetes mellitus in Cyprus. *BMC Public Health* 12, 334. doi:10.1186/1471-2458-12-334
- Malik, V.S., Popkin, B.M., Bray, G.A., Després, J.-P., Hu, F.B., 2010a. Sugar-Sweetened Beverages, Obesity, Type 2 Diabetes Mellitus, and Cardiovascular Disease Risk. *Circulation* 121, 1356–1364. doi:10.1161/CIRCULATIONAHA.109.876185
- Malik, V.S., Popkin, B.M., Bray, G.A., Després, J.-P., Willett, W.C., Hu, F.B., 2010b. Sugar-Sweetened Beverages and Risk of Metabolic Syndrome and Type 2 Diabetes. *Diabetes Care* 33, 2477–2483. doi:10.2337/dc10-1079
- Malik, V.S., Schulze, M.B., Hu, F.B., 2006. Intake of sugar-sweetened beverages and weight gain: a systematic review. *Am. J. Clin. Nutr.* 84, 274–288.

- Maull, E.A., Ahsan, H., Edwards, J., Longnecker, M.P., Navas-Acien, A., Pi, J., Silbergeld, E.K., Styblo, M., Tseng, C.-H., Thayer, K.A., Loomis, D., 2012. Evaluation of the Association between Arsenic and Diabetes: A National Toxicology Program Workshop Review. *Environ. Health Perspect* 120, 1658–1670. doi:10.1289/ehp.1104579
- Meliker, J.R., Wahl, R.L., Cameron, L.L., Nriagu, J.O., 2007. Arsenic in drinking water and cerebrovascular disease, diabetes mellitus, and kidney disease in Michigan: a standardized mortality ratio analysis. *Environ Health* 6, 4. doi:10.1186/1476-069X-6-4
- Monosson, E., 2005. Chemical Mixtures: Considering the Evolution of Toxicology and Chemical Assessment. *Environ. Health Perspect.* 113, 383–390. doi:10.1289/ehp.6987
- Moon, K., Guallar, E., Navas-Acien, A., 2012. Arsenic exposure and cardiovascular disease: an updated systematic review. *Current atherosclerosis reports* 14, 542–555.
- Moon, K.A., Guallar, E., Umans, J.G., Devereux, R.B., Best, L.G., Francesconi, K.A., Goessler, W., Pollak, J., Silbergeld, E.K., Howard, B.V., et al., 2013. Association between exposure to low to moderate arsenic levels and incident cardiovascular disease: a prospective cohort study. *Ann. Intern. Med.* 159, 649–659.
- Mordukhovich, I., Wright, R.O., Amarasiriwardena, C., Baja, E., Baccarelli, A., Suh, H., Sparrow, D., Vokonas, P., Schwartz, J., 2009. Association between low-level environmental arsenic exposure and QT interval duration in a general population study. *Am. J. Epidemiol.* 120, 98–140.
- Mordukhovich, I., Wright, R.O., Hu, H., Amarasiriwardena, C., Baccarelli, A., Litonjua, A., Sparrow, D., Vokonas, P., Schwartz, J., 2012. Associations of toenail arsenic, cadmium, mercury, manganese, and lead with blood pressure in the normative aging study. *Environ. Health Perspect.* 120, 98.
- Murphy, E.A., Post, G.B., Buckley, B.T., Lippincott, R.L., Robson, M.G., 2012. Future Challenges to Protecting Public Health from Drinking-Water Contaminants. *Annu. Rev. Public Health* 33, 209–224. doi:10.1146/annurev-publhealth-031811-124506
- Navas-Acien, A., Silbergeld, E.K., Streeter, R.A., Clark, J.M., Burke, T.A., Guallar, E., 2006. Arsenic Exposure and Type 2 Diabetes: A Systematic Review of the Experimental and Epidemiologic Evidence. *Environ. Health Perspect.* 114, 641–648. doi:10.1289/ehp.8551
- Nazzal, Y., Zaidi, F.K., Ahmed, I., Ghrefat, H., Naeem, M., Al-Arifi, N.S.N., Al-Shaltoni, S.A., Al-Kahtany, K.M., 2015. The combination of principal component analysis and geostatistics as a technique in assessment of groundwater hydrochemistry in arid environment. *Curr. Sci. (00113891)* 108, 1138–1145.
- O'Connor, A.M., Sargeant, J.M., 2014. Meta-analyses including data from observational studies. *Preventive veterinary medicine* 113, 313–322.
- Onufrak, S.J., Park, S., Sharkey, J.R., Sherry, B., 2014. The relationship of perceptions of tap water safety with intake of sugar-sweetened beverages and plain water among US adults. *Public Health Nutr.* 17, 179–185. doi:10.1017/S1368980012004600
- Peters, S.C., Blum, J.D., Klaue, B., Karagas, M.R., 1999. Arsenic occurrence in New Hampshire drinking water. *Environ. Sci. Technol.* 33, 1328–1333.

- Peterson, H., Torchia, M., 2008. Safe drinking water for rural Canadians. *Can. Med. Assoc. J.* 179, 55–55. doi:10.1503/cmaj.1080061
- Public Health Agency of Canada, 2014. Economic Burden of Illness in Canada, 2005-2008. URL <http://www.phac-aspc.gc.ca/publicat/ebic-femc/2005-2008/assets/pdf/ebic-femc-2005-2008-eng.pdf>.
- Rahman, M., Tondel, M., Ahmad, S.A., Axelson, O., 1998. Diabetes mellitus associated with arsenic exposure in Bangladesh. *American Journal of epidemiology* 148, 198–203.
- Renn, O., Burns, W.J., Kasperson, J.X., Kasperson, R.E., Slovic, P., 1992. The social amplification of risk: Theoretical foundations and empirical applications. *J. Soc. Issues* 48, 137–160.
- Rothman, K.J., 1981. Induction and Latent Periods. *Am. J. Epidemiol.* 114, 253–259.
- Sánchez-Martos, F., Jiménez-Espinosa, R., Pulido-Bosch, A., 2001. Mapping groundwater quality variables using PCA and geostatistics: a case study of Bajo Andarax, southeastern Spain. *Hydrol. Sci. J.* 46, 227–242. doi:10.1080/02626660109492818
- Satyaji Rao, Y.R., Keshari, A.K., Gosain, A.K., 2009. Evaluation of regional groundwater quality using PCA and geostatistics in the urban coastal aquifer, East Coast of India. *Int. J. Environ. Waste Manag.* 5, 163–180.
- Saylor, A., Prokopy, L.S., Amberg, S., 2011. What's wrong with the tap? Examining perceptions of tap water and bottled water at Purdue University. *Environ. Manage.* 48, 588–601.
- Schulze, M.B., 2004. Sugar-Sweetened Beverages, Weight Gain, and Incidence of Type 2 Diabetes in Young and Middle-Aged Women. *JAMA J. Am. Med. Assoc.* 292, 927–934. doi:10.1001/jama.292.8.927
- Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., Von Gunten, U., Wehrli, B., 2006. The challenge of micropollutants in aquatic systems. *Science* 313, 1072–1077.
- Serdar, M., Bakir, F., Hasimi, A., Celik, T., Akin, O., Kenar, L., Aykut, O., Yildirimkaya, M., 2009. Trace and toxic element patterns in nonsmoker patients with noninsulin-dependent diabetes mellitus, impaired glucose tolerance, and fasting glucose. *International journal of diabetes in developing countries* 29, 35-40.
- Shyu, G.-S., Cheng, B.-Y., Chiang, C.-T., Yao, P.-H., Chang, T.-K., 2011. Applying Factor Analysis Combined with Kriging and Information Entropy Theory for Mapping and Evaluating the Stability of Groundwater Quality Variation in Taiwan. *Int. J. Environ. Res. Public Health* 8, 1084–1109. doi:10.3390/ijerph8041084
- Sketchell, J., Shaheen, N., 2000. Ground water quality in rural Saskatchewan—Emerging issues for drinking water, in: *Maintaining Drinking Water Quality—Lessons from the Prairies and Beyond. Proceedings of the 9th National Conference on Drinking Water.* Regina. Saskatchewan. Canada. (Ed. W. Robertson.). pp. 242–258.
- Slovic, P., Fischhoff, B., Lichtenstein, S., 1982. Why study risk perception? *Risk Anal.* 2, 83–93.

- Smith, A.H., Steinmaus, C.M., 2009. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. *Annu. Rev. Public Health* 30, 107–122.  
doi:10.1146/annurev.publhealth.031308.100143
- Statistics Canada, 2011. Population, urban and rural, by province and territory (Saskatchewan). URL <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo62i-eng.htm>.
- Statistics Canada, 2011. Households and the Environment: Analysis. URL <http://www.statcan.gc.ca/pub/11-526-x/2013001/part-partie1-eng.htm>.
- Stookey, J.D., Constant, F., Popkin, B.M., Gardner, C.D., 2008. Drinking Water Is Associated With Weight Loss in Overweight Dieting Women Independent of Diet and Activity. *Obesity* 16, 2481–2488. doi:10.1038/oby.2008.409
- Thompson, T.S., 2003. General Chemical Water Quality of Private Groundwater Supplies in Saskatchewan, Canada. *Bull. Environ. Contam. Toxicol.* 70, 0447–0454.  
doi:10.1007/s00128-003-0007-3
- Thompson, T.S., 2001. Nitrate concentrations in private rural drinking water supplies in Saskatchewan, Canada. *Bull. Environ. Contam. Toxicol.* 66, 64–70.
- Thompson, T.S., Le, M.D., Kasick, A.R., Macaulay, T.J., 1999. Arsenic in Well Water Supplies in Saskatchewan. *Bull. Environ. Contam. Toxicol.* 63, 478–483.  
doi:10.1007/s001289901005
- Tsai, S.-M., Wang, T.-N., Ko, Y.-C., 1999. Mortality for certain diseases in areas with high levels of arsenic in drinking water. *Archives of Environmental Health: An International Journal* 54, 186–193.
- Tseng, C.-H., 2004. The potential biological mechanisms of arsenic-induced diabetes mellitus. *Toxicol. Appl. Pharmacol.* 197, 67–83. doi:10.1016/j.taap.2004.02.009
- Tsuji, J.S., Perez, V., Garry, M.R., Alexander, D.D., 2014. Association of low-level arsenic exposure in drinking water with cardiovascular disease: A systematic review and risk assessment. *Toxicology* 323, 78–94. doi:10.1016/j.tox.2014.06.008
- Turgeon, S., Rodriguez, M.J., Thériault, M., Levallois, P., 2004. Perception of drinking water in the Quebec City region (Canada): the influence of water quality and consumer location in the distribution system. *J. Environ. Manage.* 70, 363–373.  
doi:10.1016/j.jenvman.2003.12.014
- Villanueva, C.M., Kogevinas, M., Cordier, S., Templeton, M.R., Vermeulen, R., Nuckols, J.R., Nieuwenhuijsen, M.J., Levallois, P., 2014. Assessing exposure and health consequences of chemicals in drinking water: current state of knowledge and research needs. *Environ. Health Perspect. Online* 122, 213.
- Wang, S.-L., Chiou, J.-M., Chen, C.-J., Tseng, C.-H., Chou, W.-L., Wang, C.-C., Wu, T.-N., Chang, L.W., 2003. Prevalence of non-insulin-dependent diabetes mellitus and related vascular diseases in southwestern arseniasis-endemic and nonendemic areas in Taiwan. *Environ Health Perspect* 111, 155–159.

- Wang, W., Xie, Z., Lin, Y., Zhang, D., 2014. Association of inorganic arsenic exposure with type 2 diabetes mellitus: a meta-analysis. *J. Epidemiol. Community Health* 68, 176–184. doi:10.1136/jech-2013-203114
- Water Security Agency, 2016. Municipal Drinking Water Quality Monitoring Guidelines. URL <http://www.saskh2o.ca/pdf/epb202.pdf>
- Water Security Agency, n.d. Saskatchewan's Drinking Water Quality Standards and Objectives (Summarized). URL <http://www.saskh2o.ca/pdf/epb507.pdf>
- Wu, M.-M., Chiou, H.-Y., Hsueh, Y.-M., Hong, C.-T., Su, C.-L., Chang, S.-F., Huang, W.-L., Wang, H.-T., Wang, Y.-H., Hsieh, Y.-C., others, 2006. Effect of plasma homocysteine level and urinary monomethylarsonic acid on the risk of arsenic-associated carotid atherosclerosis. *Toxicology and applied pharmacology* 216, 168–175.

## **CHAPTER 2: RISK FACTORS ASSOCIATED WITH PERCEPTIONS OF DRINKING WATER QUALITY IN RURAL SASKATCHEWAN**

This is the authors accepted manuscript of an article published as the version of record in Canadian Water Resource Journal 13th January 2015: <http://www.tandfonline.com/http://dx.doi.org/10.1080/07011784.2014.985513>

McLeod, L., Bharadwaj, L., Waldner, C., 2015. Risk factors associated with perceptions of drinking water quality in rural Saskatchewan. Canadian Water Resources Journal / Revue canadienne des ressources hydriques 40, 36–46. doi:10.1080/07011784.2014.985513

This manuscript reproduced under licence with minor modifications for formatting. My contribution to the manuscript was design of the sampling strategy, data collection and management, statistical analysis, and writing the manuscript.

*This chapter established a baseline for understanding the perceptions of water quality and risks to health from drinking water among rural Saskatchewan residents. An innovative method was used to distribute questionnaires to a large sample of rural Saskatchewan residents, and the results were analyzed using methods to account for clustering resulting from the sampling strategy. A better understanding of the perceptions of rural Saskatchewan residents is important for the development of public health education and programs designed to minimize risks associated with drinking water supplies in rural areas. Understanding these perceptions and how they influence behaviors that modify exposure to contaminants in drinking water is also important for risk assessments and epidemiologic studies where exposure assessment could be refined based on personal behavior.*

## **2.1 Abstract**

A cross-sectional study used anonymous postal questionnaires to investigate risk factors associated with perceptions of water quality and risk from drinking water in rural Saskatchewan. From the 7500 questionnaires mailed out, the 2065 returned responses were analyzed using generalized linear mixed models. Reporting a drinking water advisory significantly increased the likelihood of any aesthetic complaint with tap water. Using truck-delivered water, being older, being male and living in an area for more than 10 years significantly decreased the likelihood of an aesthetic complaint. Having an aesthetic complaint significantly increased the likelihood of believing that tap water is not safe to drink. However, using a community water supply decreased this likelihood. Reporting a water advisory significantly increased the likelihood of believing tap water was not safe, but the magnitude of the effect was greater for those who used a groundwater source than those who did not. Using a groundwater source significantly decreased the likelihood of believing the tap water was not safe, but only for those who never had a water advisory.

## **2.2 Introduction**

Perceptions of drinking water quality and risk are influenced by a complex set of factors relating to sensory perception, risk tolerance and social, psychological and economic factors. In the context of drinking water, risk perception has been defined as “an individual’s subjective judgment (based on aesthetic and non-aesthetic qualities) about drinking water” (Anadu and Harding, 2000). Risk perception largely depends on intuitive judgment and is also influenced by cultural and social factors (Slovic, 1987), including race, gender and socioeconomic factors (Finucane et al., 2000; Dosman et al., 2001).

In a review of the factors related to water quality and risk perception, Doria (2010) identified several factors that influence these perceptions, including aesthetic qualities, prior experience, familiarity, interpersonal information, trust and control issues, and personal and cultural influences. Studies of water quality and risk perception in Canada and elsewhere have shown that the aesthetic qualities of drinking water, especially taste and odor, are associated with quality and risk perception (Jones et al., 2007; Doria et al., 2009). However, many potentially



harmful water contaminants are not expected to change the aesthetic qualities of water, and aesthetic characteristics can be unreliable for making risk judgments (Turgeon et al., 2004). Nevertheless, it has been suggested that personal experiences typically have the largest impact on perceptions of drinking water risk and quality. The aesthetic qualities of the water represent the main experience most consumers have with their water quality and are, therefore, the most readily available metric for making judgments (Doria, 2010).

A sense of control over a hazard could reduce the risk perceived (Doria, 2010). Residents with a private water supply could feel they have greater control over their water supply and therefore have more confidence in its safety. Trust in the water supplier, relevant for those with community treated water supplies, has also been identified as important in the perception of water quality and risk (Doria et al., 2009; McSpirit and Reid, 2011; Saylor et al., 2011), as has familiarity with the water supply (Dietrich, 2006; Doria, 2010). Demographic and socioeconomic factors have also been found to play a role in water quality perceptions, although their relative importance and effect vary depending on the community (Turgeon et al., 2004; Doria, 2010; McSpirit and Reid, 2011).

An understanding of the factors associated with water quality and risk perception is important for developing public health strategies that promote safe drinking water. This is particularly important in rural areas, where a number of water sources of varied quality might be employed (Corkal et al., 2011), private sources are common and effective dissemination of timely information and educational materials can be challenging. Furthermore, we are unaware of any large-scale studies of water quality and risk perceptions in rural Saskatchewan, where the types of water sources used can vary considerably.

Differences in survey methods used across communities can make comparing studies of water quality and risk perception challenging. Insufficient clarification of the dimensions of quality and risk pose an additional barrier to the comparison of factors associated with these perceptions. Although the concepts of quality and risk might be expected to be closely related, Janmaat (2007) suggested that concerns about quality and risk are somewhat separable, with concerns tending to be strongest in only one of these dimensions. Our primary objectives were to investigate risk factors associated with reporting any aesthetic complaints about the household's

tap water, and also with the perception that the tap water is not safe to drink. The secondary objectives were to examine risk factors associated with the fear that the household's water source will become contaminated, and the perception that someone had become ill as a result of drinking the household's tap water.

## **2.3 Materials and methods**

### *2.3.1 Design*

A postal questionnaire was sent to 7500 rural households in six regions of Saskatchewan in the fall of 2011. The questionnaire was anonymous and distributed through Canada Post's Unaddressed AdMail service. Postal code geography files (Platinum Postal Code Suite 2006; DMTI Spatial Inc., Markham, ON) were used in conjunction with Canada Post Householder Counts to select postal codes to which the questionnaire was distributed. A commercial geographic information system (ArcMAP, ESRI, Redlands, CA) was used to calculate the smallest radius for each region that would be necessary to include centroids of eligible postal codes encompassing 1250 houses and farms. Postal codes with no farms were excluded, and for postal codes that contained more than 200 houses, the survey was sent only to farms to ensure that the distribution of the questionnaire would be primarily to rural households. Questionnaires were sent to 1250 households from between nine and 12 postal codes (median = 10) in each of the six regions, for a total of 60 postal codes. The resultant data were hierarchical with clustering by postal code nested within region.

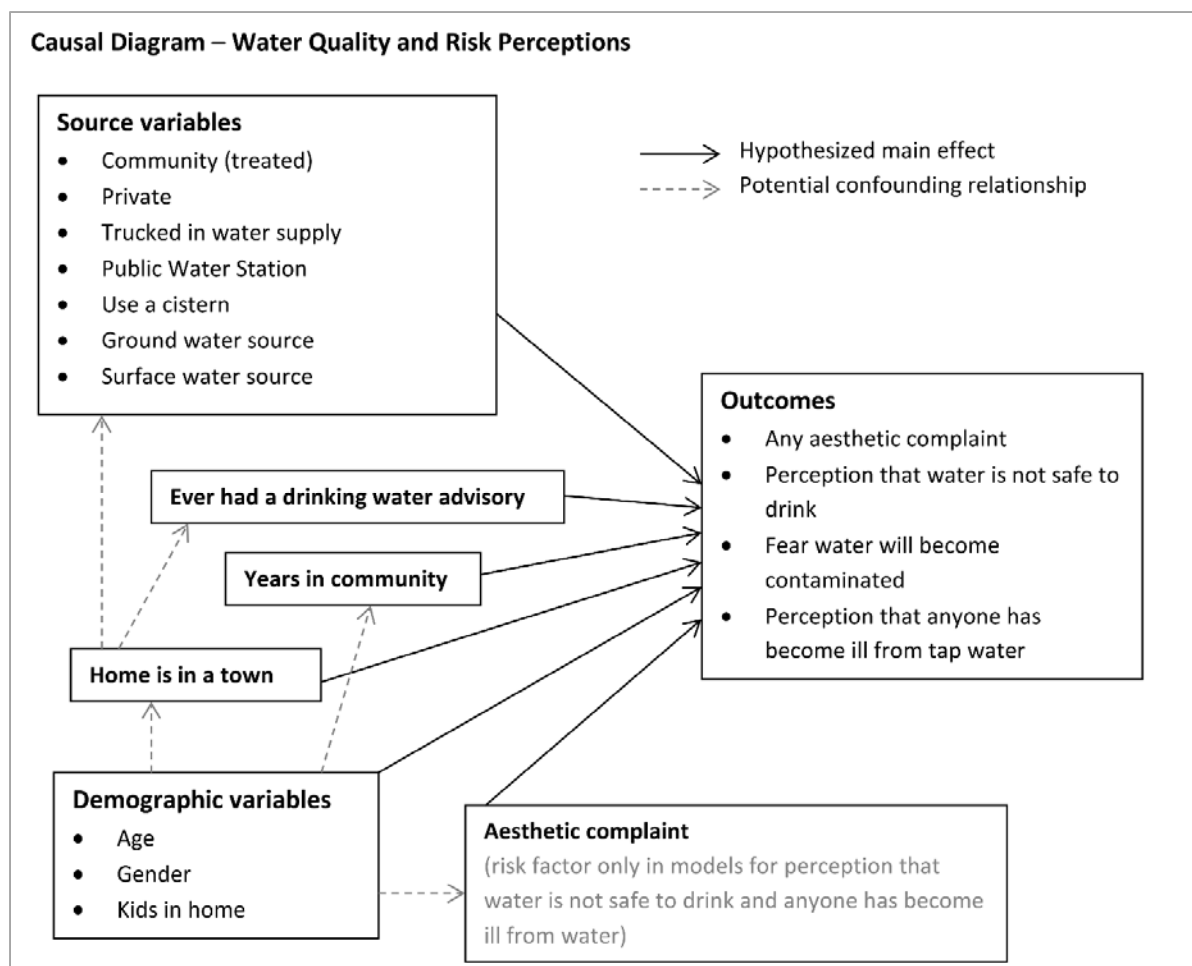
The questionnaire was four pages long and consisted of questions about water sources, perceptions of water quality and risk, experiences with water advisories and drinking water choices. It was based on a questionnaire that had been piloted in a different community the previous year. We asked that one member of the household over the age of 18 fill out the questionnaire and return it in a postage-paid, pre-addressed envelope. Because the questionnaires were not addressed to specific households, distributing reminders was not practical. The informed consent of the participants was obtained, and the study protocol was approved by the Behavioral Research Ethics Board of the University of Saskatchewan (Beh 10-184).

### *2.3.2 Outcomes and risk factors*

The primary outcomes of interest in this analysis were whether residents specified any aesthetic complaint about their household's tap water (odor, bad taste, discoloration or cloudiness), and whether residents perceived that their household tap water was not safe to drink. Fear that the household's water source will become contaminated and the perception that someone had become ill from drinking the household's tap water were also modeled as outcomes.

Reporting an aesthetic complaint was a dichotomous variable created from a question which allowed multiple responses regarding characteristics of the household's tap water. If any of odor, bad taste, discoloration or cloudiness were selected, the aesthetic complaint variable was coded as "yes," and if none were selected, the variable was coded as "no." The frequencies of the individual aesthetic complaints were also calculated, and the agreement among pairwise combinations was estimated using Kappa coefficients ( $\kappa$ ).

A causal diagram (Figure 2.1) was constructed to help guide model development. The risk factors included a variety of measures related to household tap water sources: use of a community treated water supply, use of a private water supply, water delivered by truck, use of a public water station, whether the household had a cistern and whether the water source was groundwater or surface water. In rural areas, households can use more than one water source; therefore, each of the possible types of water supply was analyzed separately. Water sources were not mutually exclusive.



**Figure 2.1** Causal diagram used to guide model development for each of the outcomes

History of having had any type of drinking water advisory, past or present, in the current household was assessed as a risk factor. As well, whether the home was in a town, length of time residing in the current community, age, gender and whether there were children under 18 years old residing in the household were also analyzed as risk factors.

Additionally, because aesthetic characteristics have been linked to drinking water risk perception in the literature, having an aesthetic complaint was included as a risk factor in the models both for the perception that the household tap water is not safe to drink and for the perception that someone has become ill from drinking the tap water.

Six age categories were recorded on the questionnaire. The three youngest age groups were collapsed into a single category due to low numbers of responses in these categories, so that only

four categories were used in the analysis. Four possible categories for the number of years residing in the community were also collapsed into two categories for analysis.

### 2.3.3 Statistical analysis

Each outcome was modeled using a generalized linear mixed model (GLMM), specifying a binomial distribution and logit link function. When exploration of differences between groups is not the primary goal of the analysis, these models parsimoniously account for clustering by partitioning the overall variance in the data into variance components. Random intercepts were included for both postal code (pc) and region (reg) to account for any clustering in the data arising from the hierarchical structure of the data (Equation 2.1):

$$\text{logit}(p_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots \beta_k X_k + \mu_{\text{pc}(i)} + \nu_{\text{reg}(i)} \quad (2.1)$$

$$\begin{aligned} p_i &= \Pr(y_i = 1) \\ \mu_{\text{pc}(i)} &\sim N(0, \sigma_{\text{pc}}^2) \\ \nu_{\text{reg}(i)} &\sim N(0, \sigma_{\text{reg}}^2) \end{aligned}$$

Models were built for each outcome by first screening each risk factor individually in a logistic mixed effects model with random intercepts for postal code and region; any risk factor with a p-value  $< 0.2$  was retained for consideration when building the final model. Prior to considering all remaining risk factors ( $p < 0.2$ ), an intermediate model was used to identify the best subset of retained risk factors related to water source, keeping only those with a p-value  $< 0.05$ . Manual backwards selection was used to build the final main-effects model, retaining only risk factors with a p-value  $< 0.05$ . The last risk factor removed from the model was assessed for confounding based on whether its inclusion in the model led to a change greater than 10% in the regression coefficients for other risk factors. If age, gender and/or children in the home were not retained in the model but were unconditionally associated with the outcome, they were assessed for confounding by the same criteria. Biologically plausible interactions between risk factors retained in the final model were assessed at a 0.05 level of significance; in the case of categorical variables, a type 3 likelihood ratio test was used to determine if the interaction was significant. Akaike information criterion (AIC) values were used to assess competing models where there was doubt regarding the best model fit.

Models were built in Stata (StataCorp LP, College Station, Texas) with the `xtmelogit` command using a Laplacian approximation for efficiency. Using the risk factors identified in the model-building process, the final model parameters were estimated with the user-written Stata program `gllamm` (Rabe-Hesketh et al., 2004) using adaptive quadrature with 12 integration points. Population averaged probabilities were estimated using the `gllapred` marginal function (Rabe-Hesketh and Skrondal, 2008).

Odds ratios were obtained by exponentiating the regression coefficients ( $e^{\beta}$ ) for the risk factors included in each model. Odds ratios represent the relative odds of the occurrence of an outcome in those with a risk factor compared to those without (or with a different level of the risk factor). Odds ratios (OR) were reported along with corresponding 95% confidence intervals (CI).

The proportion of the variance accounted for by postal code (Equation 2.2) and region (Equation 2.3) was examined for each of the outcomes, using an approximation of the variance partition coefficient (VPC) for the binomial outcome based on the latent response variable model (Browne et al., 2005):

$$VPC = \frac{\sigma_{pc}^2}{\sigma_{reg}^2 + \sigma_{pc}^2 + \frac{\pi^2}{3}} \quad (2.2)$$

$$VPC = \frac{\sigma_{reg}^2}{\sigma_{reg}^2 + \sigma_{pc}^2 + \frac{\pi^2}{3}} \quad (2.3)$$

Values were missing for all outcomes and risk factors from at least one survey. Any observations that were missing values for any of the risk factors or outcome for a given model were excluded from analyses including that variable. Therefore, the final number of observations used in each model varied and is reported for each model. Model assumptions were examined by evaluating the distribution of the residuals at each random effects level using Q–Q plots. Residuals were also examined for the presence of outliers and influential data points by plotting the standardized residuals at each of the random effects levels.

## **2.4 Results**

Of the 7500 questionnaires mailed out, 2074 were returned. One was excluded because it was not filled out, one was discarded because it was received subsequent to data entry and analysis, and seven were excluded from the analysis because the postal code identifier had been removed from the questionnaire before returning the questionnaire. As a result, 2065 total observations were available for use in this analysis, for an effective response rate of 27.5%.

The median number of responses per region was 353 (range 327–368), while the median number of responses per postal code was 44 (range 2–108); frequencies of each outcome and risk factor were calculated (Table 2.1).

**Table 2.1** Total numbers of complete and missing responses and the proportion of respondents at each level for each of the outcomes modeled and risk factors evaluated

	Complete	Missing	Response	Frequency	
	<i>n</i>	<i>n</i>		<i>n</i>	% <sup>1</sup>
<b>Outcomes</b>					
Any aesthetic complaint about tap water	1984	81	Yes	501	25.3
			No	1483	74.8
Believed tap water not safe to drink	1984	81	Yes	235	11.8
			No	1749	88.2
Fear tap water will become contaminated	1988	77	Yes	706	35.5
			No	1282	64.5
Someone had been ill from tap water	1784	281	Yes	57	3.2
			No	1727	96.8
<b>Risk Factors</b>					
Private water supply	2059	6	Yes	1249	60.7
			No	810	39.3
Community (treated) water supply	2059	6	Yes	640	31.1
			No	1419	68.9
Water delivered by truck	2056	9	Yes	119	5.8
			No	1937	94.2
Used a public water station	2059	6	Yes	121	5.9
			No	1938	94.1
Had a cistern	2027	38	Yes	292	14.4
			No	1735	85.6
Ground water source	1857	208	Yes	1349	72.6
			No	508	27.4
Surface water source	1856	209	Yes	613	33.0
			No	1243	67.0
Ever had a water advisory	1981	84	Yes	485	24.5
			No	1496	75.5
Home is in a town	2047	18	Yes	525	25.6
			No	1522	74.4
Gender	2005	60	Female	1053 <sup>2</sup>	52.5
			Male	952	47.5
Age	2050	15	< 45 years	317 <sup>2</sup>	15.5
			45-54 years	446	21.8
			55-64 years	614	30.0
			≥ 65 years	673	32.8
Number of years in community	2046	19	≤10 years	403 <sup>2</sup>	19.7
			> 10 years	1643	80.3
Children reside in home	1932	133	Yes	437	22.6
			No	1495	77.4

<sup>1</sup>Percentage based on number of complete observations. <sup>2</sup>Reference category



The distributions of age categories and gender in our sample population were compared to data from the Canada 2011 Census of Population (Statistics Canada, 2011) for the Census Subdivisions included in our survey regions (Table 2.2) to assess the representativeness of our sample. Awareness of issues such as agricultural runoff or pollution affecting the household water source was reported by 21.8% of respondents.

**Table 2.2** Comparison of key demographic variables between the sample population and the Statistics Canada 2011 Census of Population for the rural Census Subdivisions included within the survey regions.

<b>Category</b>	<b>Survey respondents<sup>1</sup></b>		<b>2011 Census of Population (%)</b>
	<b>N</b>	<b>(%)</b>	
Female	1053	52.5	47.3
Male	952	47.5	52.7
18-45 years	317	15.5	35.7
45-54 years	446	21.8	23.2
55-65 years	614	30.0	21.7
≥65 years	673	32.8	19.3

<sup>1</sup>Total number of respondents: 2005 for gender and 2050 for age

#### *2.4.1 Having any aesthetic complaint*

Overall, 501 (25.3%) respondents were dissatisfied with at least one of taste, odor, color or cloudiness of their tap water, and were classified as having any aesthetic complaint about their tap water. Of the respondents who knew the source of their tap water, the proportion with any aesthetic complaint was similar for those using groundwater sources (23.8%) and those using surface water sources (24.0%). The frequencies of each type of complaint were also broken down by type of water source used (Table 2.3). The types of aesthetic complaint were not mutually exclusive and of the 501 respondents with any complaint, 236 (47.1%) specified more than one type of aesthetic complaint. The agreement between the various aesthetic concerns ranged from a Kappa statistic of 0.19 (between taste and cloudiness) to 0.48 (taste and odor) and indicated slight to moderate agreement between the various types of complaints over all respondents (Dohoo et al., 2012). In the open comments area of the survey, 159 respondents

(8.0%) wrote that their tap water had a high mineral content or hardness; of these, 141 (88.7%) used a groundwater source.

**Table 2.3** Proportion of respondents identifying each type of aesthetic complaint overall and by the type of water source used.

Complaint	% of all respondents (n = 1984)	By type of water source used					
		Ground water (n = 1293)	Surface water (n = 592)	Private supply (n=1194)	Community supply (n=620)	Delivered by Truck (n=114)	Public Water Station (n=119)
Odor	12.5%	10.5%	13.0%	12.3%	13.7%	7.9%	14.3%
Taste	14.5%	13.1%	13.7%	11.5%	20.3%	5.3%	19.3%
Discolored	11.2%	11.1%	10.1%	13.4%	7.7%	2.6%	13.4%
Cloudy	4.8%	3.5%	6.6%	4.8%	4.0%	2.6%	5.9%

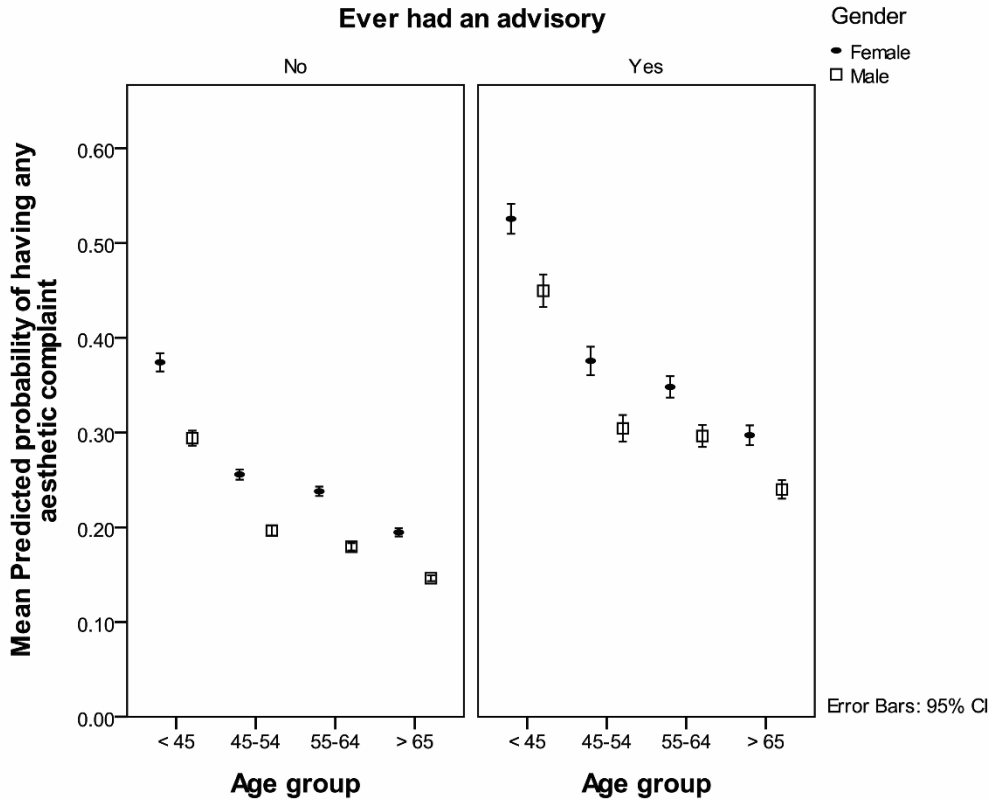
After accounting for other significant risk factors (Table 2.4), respondents who had their water delivered by truck were less likely to have an aesthetic complaint ( $p = 0.02$ ) than those who did not. Those who reported any type of water advisory were more likely to report an aesthetic concern ( $p < 0.001$ ) (Figure 2.2) than those who had not experienced an advisory. Compared to the youngest age group, respondents in each of the older age categories were less likely to have an aesthetic complaint ( $p \leq 0.006$ ) (Figure 2.2). Similarly, those over 65 were less likely to report aesthetic concerns than those 45–54 (OR 0.7, 95% CI 0.5–1.0,  $p = 0.02$ ). Being male ( $p = 0.005$ ; Figure 2.2) and having lived in the community for longer than 10 years ( $p = 0.007$ ) were also associated with a decreased likelihood of having an aesthetic complaint.

**Table 2.4** Risk factors significantly associated with having any type of aesthetic complaint with the household's tap water in the final multivariable model.

<b>Risk Factor</b>	<b>OR</b>	<b>95% CI</b>	<b>p value</b>
Trucked water delivery	0.5	0.3, 0.9	0.02
Ever had a water advisory	1.8	1.4, 2.3	< 0.001
Male	0.7	0.6, 0.9	0.005
Female	REF		
Age			< 0.001 <sup>1</sup>
Age < 45	REF		
Age 45-54	0.6	0.4, 0.9	0.006
Age 55-64	0.6	0.4, 0.8	< 0.001
Age 65+	0.4	0.3, 0.6	< 0.001
≤ 10 years in community	REF		
>10 Years in community	0.7	0.5, 0.9	0.007
<b>Random Effects</b>	<b>Variance</b>	<b>SE</b>	
Postal Code	0.140	0.073	
Region	0.034	0.040	
Number of observations = 1865			

OR: odds ratio. CI: confidence interval. REF: reference category. SE: standard error.

<sup>1</sup>Wald Type III test of fixed effect for age



**Figure 2.2** Predicted probabilities of having any aesthetic complaint, comparing those who had a drinking water advisory to those who did not, for men and women in each age category averaged over the use of water delivered by truck and length of time residing in community. CI: confidence interval.

Because the effect of age on taste satisfaction varied by community in a previous study (Turgeon et al., 2004), we tested if the effect of age varied by region, and found no evidence that this was the case in our study.

The proportion of variance accounted for by postal code (4.1%) was greater than the proportion accounted for by region (1.0%). This represented a small change compared to the random effects from the null model (in which postal code accounted for 6.2% of the total variance, and region 1.0%), suggesting that the fixed effects did not account for much of the between-group variation in having an aesthetic complaint.

#### *2.4.2 The perception that tap water is not safe to drink*

After accounting for other risk factors (Table 2.5), those who reported having any aesthetic complaint about the tap water were more likely to agree that their tap water was not safe to drink compared to those who did not report any aesthetic complaints ( $p < 0.001$ ; Figure 2.3).

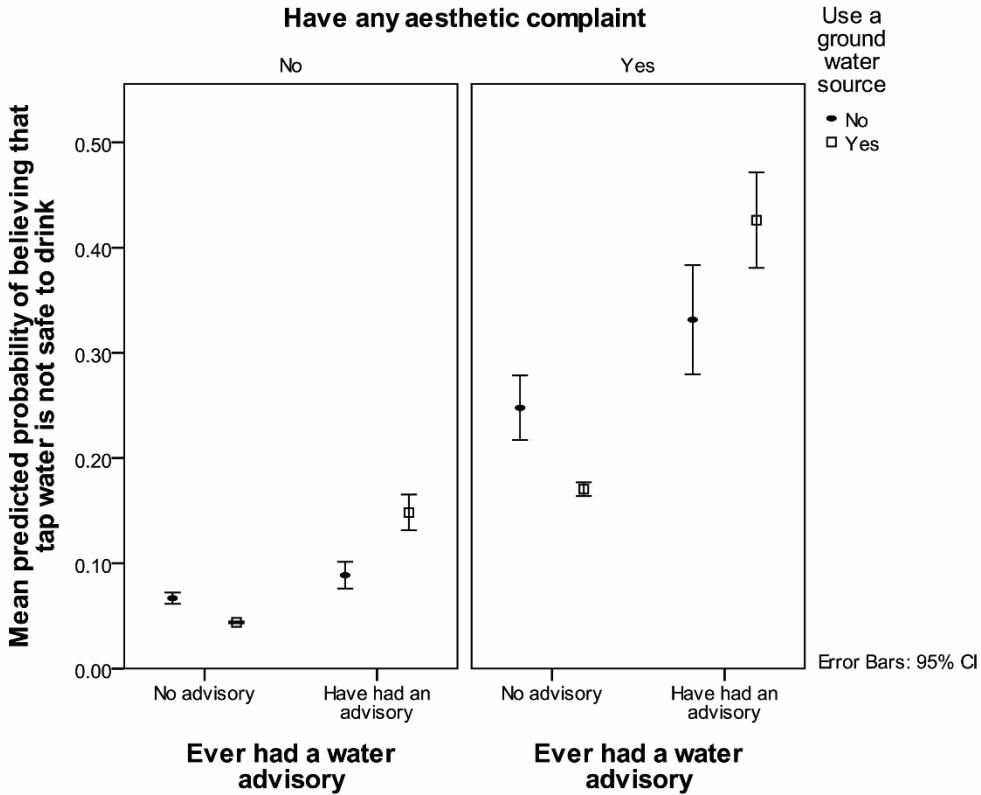
Respondents who used a community water supply were less likely to believe that their tap water was not safe when compared to not using a community supply ( $p < 0.001$ ). An interaction was identified between the use of a groundwater source and having experienced any type of drinking water advisory. Compared to those who did not use a groundwater source, those who used a groundwater source were less likely to believe that the household tap water was not safe to drink, but only when they had never had a drinking water advisory ( $p = 0.001$ ; Figure 2.3). Compared to those who reported never having had a drinking water advisory, respondents who had experienced a drinking water advisory were more likely to believe that their tap water was not safe to drink, but the magnitude of this effect was greater for the respondents who used a groundwater source ( $OR = 7.3$ ,  $p < 0.001$ ) than for those who did not use a groundwater source ( $OR = 3.0$ ,  $p < 0.001$ ; Figure 2.3).

**Table 2.5** Risk factors significantly associated, in the final multivariable model, with the perception that household tap water is not safe to drink.

<b>Risk Factors</b>	<b>OR</b>	<b>95% CI</b>	<b>p value</b>
Community treated water supply	0.2	0.1, 0.3	< 0.001
Any aesthetic complaint	4.6	3.3, 6.3	< 0.001
Ground water source x ever had advisory			0.02 <sup>1</sup>
Use a groundwater source compared to not using groundwater for those who have ever had a water advisory	1.1	0.6, 2.0	0.77
Use a groundwater source compared to not using groundwater for those who have <i>not</i> had a water advisory	0.4	0.3, 0.7	0.001
Ever had an advisory compared to not having had an advisory for those who use a groundwater source	7.3	4.6, 11.5	< 0.001
Ever had an advisory compared to not having had an advisory for those who do <i>not</i> use a groundwater source	3.0	1.6, 5.6	0.001
<b>Variances of Random Effects</b>	<b>Variance</b>	<b>SE</b>	
Postal Code	0.039	0.080	
Region	0.003	0.030	
Number of observations = 1734			

OR: odds ratio. CI: confidence interval. SE: standard error.

<sup>1</sup>Type III test of fixed effect for interaction



**Figure 2.3** Predicted probabilities of believing the tap water is not safe to drink for the interaction between reporting a drinking water advisory and use of a groundwater source, separated by whether or not an aesthetic complaint about the water was identified, and averaged over the use of a community water supply. CI: confidence interval.

Grouping by postal code accounted for 1.0% of the total variance in the model for the perception that tap water is not safe, while region accounted for 0.1%. This represents an improvement over the null model (in which postal code accounted for 6.4% of the total variance, and region 0.9%), suggesting that the fixed effects in this model explain most of the differences in the perception that tap water is not safe to drink between postal codes and regions of the province.

#### 2.4.3 Fear that water will become contaminated

After accounting for other risk factors, respondents who reported having had a drinking water advisory were more likely to fear that their household's water would become contaminated compared to those who did not experience an advisory (OR = 2.2, 95% CI 1.7–2.8,  $p < 0.001$ ). Those who used a community water supply were less likely to fear their drinking water would become contaminated than those who used another source (OR = 0.5, 95% CI 0.4–0.6,  $p < 0.001$ ).

0.001). Those who had lived in the community for more than 10 years were less likely to fear contamination compared to those who had lived in the community for 10 years or less (OR = 0.8, 95% CI 0.6–1.0,  $p = 0.04$ ). Age was associated with the fear that water will become contaminated ( $p = 0.003$ ). Compared to the over-65 age group, the 45–54 age group (OR = 1.7, 95% CI 1.3–2.3,  $p < 0.001$ ) and the 55–64 age group (OR = 1.4, 95% CI 1.1–1.8,  $p = 0.02$ ) were more likely to report fear that the tap water will become contaminated. Estimates in this model were adjusted for the presence of children in the home.

#### *2.4.4 Perception that someone became ill from drinking the household's tap water*

Respondents who reported a drinking water advisory (OR = 6.0, 95% CI 3.2–11.4,  $p < 0.001$ ) were more likely to report that someone had ever become ill from drinking the household tap water than those who did not, after accounting for other risk factors. A similar association was found for those who reported any aesthetic complaint about the tap water compared to those who did not (OR = 3.7, 95% CI 2.1–6.5,  $p < 0.001$ ). Those who reported use of a community water supply were less likely to report that someone had become ill as a result of drinking the household's tap water compared to those who did not use a community supply (OR = 0.2, 95% CI 0.1–0.4,  $p < 0.001$ ). Those who reported using a private water supply were also less likely to believe that their tap water had made someone ill compared to those who did not (OR = 0.3, 95% CI 0.2–0.7,  $p = 0.005$ ).

#### *2.4.5 Summary of risk factors*

A comparison of the risk factors included in each of the final models is presented in Table 2.6. Having had a water advisory was a risk factor in all of the models, and each model also included at least one variable related to the tap water source or supply. Having an aesthetic complaint was included in both models in which it was assessed as a risk factor.



**Table 2.6** Summary of the analyses for all water quality or risk perception outcomes including the number of observations used and the risk factors associated with the final model for each outcome (effect modifications italicized).

Model	<i>n</i>	Increased odds of having outcome	Decreased odds of having outcome
Any aesthetic complaint	1865	Ever had advisory	Trucked water delivery Male Increasing age Lived in community >10 years
Tap Water not safe to drink	1734	Aesthetic complaint <i>Had a water advisory - magnitude of effect greater for groundwater users</i>	Community water supply <i>Use a groundwater source, but only for those who never had water advisory</i>
Fear water will become contaminated <sup>1</sup>	1784	Ever had a water advisory Age 45-64 more likely than ≥65	Community water supply Lived in community >10 years
Anyone ever ill from tap water	1671	Aesthetic complaint Ever had an advisory	Private water supply Community water supply

<sup>1</sup>Adjusted for whether or not children reside in home.

## 2.5 Discussion

In this study, we examined different measures of perceptions related to water quality using quantitative techniques that allowed us to evaluate the combined effect of multiple risk factors and to consider the potential for similarities in opinions within communities. This study included several different regions in rural Saskatchewan, where residents use a variety of tap water supplies and sources, and where water supplies could be impacted by varied geology and land-use activities. Although the final models for each outcome differed, there were several themes that emerged. Each model included at least one risk factor related to the type of water supply used in the household, suggesting that some aspects of a household's water supply are important in water quality and risk perception. Having had a drinking water advisory was also a common predictor of negative perceptions about water quality and related risks, as was having an aesthetic complaint in the models where it was evaluated. Age was identified as a risk factor for

having any aesthetic complaint about the tap water as well as fearing that tap water will become contaminated, and gender was a risk factor for having any aesthetic complaint.

#### *2.5.1 Aesthetic complaints*

The aesthetic qualities of water can influence our perceptions of water quality (Doria et al., 2009; Doria, 2010), as well as water consumption patterns (Levallois et al., 1999; Dupont et al., 2010; Saylor et al., 2011). We investigated potential risk factors associated with having reported any aesthetic complaint about tap water. Having had any type of water advisory, the length of time in the community and using a trucked water supply were important factors in reporting any aesthetic complaint, along with age and gender.

In the present study, 25.3% of respondents indicated some aesthetic complaint about their tap water, be it unpleasant taste, odor, discoloration or cloudiness. Previous studies have examined distinct aesthetic qualities (e.g., taste and odor separately); however, because we were interested in the presence of any aesthetic complaint, we grouped our responses about aesthetic qualities into one index variable for analysis. Just 12.5% of our respondents were dissatisfied with the odor of their tap water and 14.5% were dissatisfied with its taste, compared to a recent cross-Canada study in which 33% of participants were dissatisfied with the odor of their tap water, and 31% were dissatisfied with the taste (Dupont et al., 2010). The frequency of aesthetic complaints was similar in both groundwater and surface water users.

Of the variables related to tap water supply, only the use of a trucked water supply was significant in the final model for having an aesthetic complaint, and it decreased the likelihood of having any aesthetic complaint. Possible explanations for this effect are not clear, and could be investigated further.

To our knowledge, this is the first study to examine the potential impact of experiencing water advisories on subsequent perceptions of water quality and safety. In our analysis, having had a water advisory increased the likelihood having an aesthetic complaint about the tap water. However, the mechanism behind this association is not clear. Objectionable aesthetic qualities of water could be associated with events that lead to water advisories; for example, increased turbidity is a common trigger for advisories. However, because prior experience is an important influence on perceptions of quality and safety (Doria, 2010), it could be that experiencing an

advisory might heighten concern about water quality, and increase sensitivity to water's aesthetic characteristics.

Living in an area for a longer duration reduced the likelihood of having an aesthetic complaint compared to shorter durations of residence. Familiarity has previously been reported to be an important factor in perception of water quality (Dietrich, 2006; Doria, 2010).

Females had higher odds of reporting dissatisfaction with the aesthetic qualities of tap water, which may be related to the tendency of women to attribute higher risks to hazards than males (Finucane et al., 2000). Respondents in the youngest age category also had greater odds of reporting dissatisfaction with the aesthetic qualities of their tap water. A previous study found that age was a significant risk factor in models of taste dissatisfaction, but the direction of the effect differed by community (Turgeon et al., 2004), suggesting that the relationships between age and water quality perceptions are complex and likely influenced by other factors.

Although there was not much variance attributable to differences between postal codes, and even less attributable to region, the included risk factors did not explain much of the variation in postal code and very little of the variation in region compared to the null model. This suggests that there are other unmeasured risk factors which might explain geographical differences in the perception of aesthetic qualities of drinking water in Saskatchewan which could be examined in future studies.

#### *2.5.2 Perception that tap water is not safe*

Nearly 12% of our respondents reported that they believed their tap water is not safe to drink, comparable to a recent US study in which 15% of respondents felt their tap water was unsafe to drink (Hu et al., 2011). Having any aesthetic complaint increased the likelihood of perceiving that tap water is not safe to drink, similar to findings in other studies (Jardine et al., 1999; Jones et al., 2005, 2007; Doria et al., 2009). While it has been recognized that taste and odor should not be ignored as indicators of water quality (Jardine et al., 1999), others have pointed out that aesthetic qualities are poor indicators of safety for a variety of reasons (Turgeon et al., 2004). For example, chlorine taste has been the source of taste dissatisfaction in other studies and has been associated with risk perception in those studies (Turgeon et al., 2004; Doria et al., 2009), despite the fact that chlorine should improve the microbiological safety of water. However, in the

present study, respondents who used community treated water supplies were less likely to perceive health risks compared to those who did not use community supplies.

Having had a water advisory was identified as a risk factor for having an aesthetic complaint, and both were included as risk factors in the model for having an aesthetic complaint despite concerns of endogeneity among the risk factors. It is possible that each risk factor independently affects the perception of risk from the drinking water as well as having a confounding relationship. The perception of aesthetic qualities of drinking water has previously been identified as important in risk perception, and it is plausible that having had a water advisory would have a separate, and perhaps greater, impact on risk perception than on the presence of aesthetic complaints. This study was intended to identify important risk factors in risk perception related to drinking water among rural Saskatchewan residents, rather than estimating the exact magnitude of the relationships between the risk factors, warranting consideration of all potentially relevant risk factors.

The effect modification between use of a groundwater source and having had a water advisory hints at the complexity of the relationships among different factors that affect perceptions of water quality and risk, and could reflect an understanding of the potential for long-term contamination of groundwater sources compared to surface water.

### *2.5.3 Fear that water will become contaminated*

Although relatively few respondents felt that their water was unsafe, a larger proportion (35.5%) reported fearing that their water will become contaminated in the future. Jones et al., (2006) reported that 41% of private water supply users in a southern Ontario community shared similar concerns. It is unclear from our results whether this is a general fear, or one based on the perception that there is an immediate risk to their water supply. Having experienced a water advisory was also important, lending support to the theory that it is an important risk factor in many aspects of water quality and risk perception. Use of a community water supply decreased the likelihood of fearing contamination of the water supply, as did living in an area for greater than 10 years, which perhaps are related to the issues of trust and familiarity in risk perception (Doria, 2010). Being in the oldest age category also decreased the likelihood of reporting this

concern compared to the intermediate age groups, but not when compared to the youngest age category.

#### *2.5.4 Illness due to drinking tap water*

No attempt was made to clarify the type of illness attributed to drinking the household's tap water, nor whether drinking water had been confirmed as the cause of any illness; therefore, this outcome was considered a measure of perception of health risk from drinking tap water. Just over 3% of respondents believed that their household tap water had ever made anyone ill, similar to other results in Ontario (Jones et al., 2006) and Pennsylvania (Merkel et al., 2012). Having experienced a water advisory and having an aesthetic complaint were strongly associated with the perception that the tap water had been the source of illness, supporting the idea that prior experience with the water is an important factor in perception of risks related to drinking water (Doria, 2010).

Because the use of a community water supply and a private water supply were not mutually exclusive or the only water supply options available, both types of supply were evaluated separately as risk factors. The inclusion of both in the final model for the perception that drinking the tap water had made someone ill suggests that the use of these types of water supply was associated with more confidence in the safety of the household tap water compared to other supply types (for example, public water stations or water delivered by truck).

#### *2.5.5 Limitations*

The choice to target specific regions of the province and the limited response rate to the questionnaire could have introduced selection bias and, therefore, these results might not be generalizable across rural Saskatchewan. While respondents did include residents of up to 24% of the rural municipalities in the province, females and the over-55 age groups were over-represented in the study population compared to the population in the areas to which the survey was targeted, suggesting that the generalizability of results across the population may be limited. However, the proportion of respondents who reported aesthetic complaints and fear of water contamination is comparable to other surveys done in Canada. It unclear if our study population disproportionately represents residents with awareness of issues regarding drinking water;

however, the anonymous nature of the survey made more intensive evaluation of the characteristics of nonresponders impossible.

Our questionnaire did not include questions regarding education level or income, which limited our ability to assess the generalizability of our results. These could also potentially be important factors in the perceptions of water quality, although the importance and effect of these measures has varied in past studies (Turgeon et al., 2004; McSpirit and Reid, 2011).

Given the complexity of the factors that influence water quality and risk perception, varied interpretation of the questions by different respondents could have introduced the potential for nondifferential misclassification, and limited our capacity to identify all of the potential risk factors. Also, many of the questions that directly queried aspects of quality and risk did not specify a time frame over which respondents should report their perceptions or concerns, so it was not always clear if respondents were indicating current, recent, past or average perceptions.

## **2.6 Conclusions**

Qualitative reports have provided a basis for understanding quality and risk perception, but few other studies have attempted to quantify the risk factors associated with these perceptions. In this study, we used multivariable analysis to estimate the effects of individual risk factors as well as to evaluate differences among communities and geographic regions in the province. This study underscores the complexity of the numerous factors involved in the formation of perceptions of water quality and safety. Having experienced a water advisory was an important risk factor in all of our models of perceptions of quality and risk, providing evidence that this personal experience can impact trust in a water source. The aesthetic qualities of water were important risk factors in the models for perceived health risks. Some aspect of the type of household tap water supply was also included in each of the models, although the specific risk factor included, and its effects, were varied. Personal characteristics also had some influence on water quality and risk perception. Future work should seek to better understand the factors influencing water quality and risk perceptions, and how to address these factors to effectively implement public health education programs in rural areas about appropriate water management practices, the safety of rural water supplies and the need for regular water testing.

## 2.7 References

- Anadu, E.C., Harding, A.K., 2000. Risk perception and bottled water use. *J Am Water Works Ass* 92, 82–92.
- Browne, W.J., Subramanian, S.V., Jones, K., Goldstein, H., 2005. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168, 599–613.
- Corkal, D.R., Diaz, H., Sauchyn, D., 2011. Changing Roles in Canadian Water Management: A Case Study of Agriculture and Water in Canada’s South Saskatchewan River Basin. *International Journal of Water Resources Development* 27, 647–664. doi:10.1080/07900627.2011.621103
- Dietrich, A.M., 2006. Aesthetic issues for drinking water. *Journal of water and health* 4, 11–16.
- Dohoo, I.R., Martin, S.W., Stryhn, H., Dohoo, I.R., 2012. *Methods in Epidemiologic Research*. VER Inc., Charlottetown, P.E.I.
- Doria, M. de F., 2010. Factors influencing public perception of drinking water quality. *Water Policy* 12, 1–19. doi:10.2166/wp.2009.051
- Doria, M. de F., Pidgeon, N., Hunter, P.R., 2009. Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study. *Science of the Total Environment* 407, 5455–5464.
- Dosman, D.M., Adamowicz, W.L., Hrudey, S.E., 2001. Socioeconomic determinants of health- and food safety-related risk perceptions. *Risk Analysis* 21, 307–318. doi:10.1111/0272-4332.212113
- Dupont, D., Adamowicz, W.L., Krupnick, A., 2010. Differences in water consumption choices in Canada: the role of socio-demographics, experiences, and perceptions of health risks. *Journal of water and health* 8, 671–686.
- Finucane, M.L., Slovic, P., Mertz, C.K., Flynn, J., Satterfield, T.A., 2000. Gender, race, and perceived risk: the “white male” effect. *Health, Risk & Society* 2, 159–172. doi:10.1080/713670162
- Hu, Z., Morton, L.W., Mahler, R.L., 2011. Bottled water: United States consumers and their perceptions of water quality. *International Journal of Environmental Research and Public Health* 8, 565–578.
- Janmaat, J., 2007. Divergent Drinking Water Perceptions in the Annapolis Valley. *Canadian Water Resources Journal* 32, 99–110. doi:10.4296/cwrj3202099
- Jardine, C.G., Gibson, N., Hrudey, S.E., 1999. Detection of odour and health risk perception of drinking water. *Water Science and Technology* 40, 91–98.
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Mathews, E., Carr, D.J., Henson, S.J., 2006. Public perceptions of drinking water: a postal survey of residents with private water supplies. *BMC Public Health* 6, 94–104. doi:10.1186/1471-2458-6-94

- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Henson, S.J., Mathews, E., 2007. A qualitative exploration of the public perception of municipal drinking water. *Water Policy* 9, 425–438. doi:10.2166/wp.2007.019
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Henson, S.J., Mathews, E., 2005. Public perception of drinking water from private water supplies: focus group analyses. *BMC Public Health* 5, 129. doi:10.1186/1471-2458-5-129
- Levallois, P., Grondin, J., Gingras, S., 1999. Evaluation of consumer attitudes on taste and tap water alternatives in Quebec. *Water Science and Technology* 40, 135–139.
- McSpirit, S., Reid, C., 2011. Residents' Perceptions of Tap Water and Decisions to Purchase Bottled Water: A Survey Analysis from the Appalachian, Big Sandy Coal Mining Region of West Virginia. *Society and Natural Resources* 24, 511–520.
- Merkel, L., Bicking, C., Sekhar, D., 2012. Parents' Perceptions of Water Safety and Quality. *Journal of Community Health* 37, 195–201. doi:10.1007/s10900-011-9436-9
- Rabe-Hesketh, S., Skrondal, A., 2008. *Multilevel and Longitudinal Modelling Using Stata*. Stata Press.
- Rabe-Hesketh, S., Skrondal, A., Pickles, A., 2004. Generalized multilevel structural equation modeling. *Psychometrika* 69, 167–190.
- Saylor, A., Prokopy, L.S., Amberg, S., 2011. What's wrong with the tap? Examining perceptions of tap water and bottled water at Purdue University. *Environmental management* 48, 588–601.
- Slovic, P., 1987. Perception of risk. *Science* 236, 280–285.
- Statistics Canada. 2011. *Census of Canada, 2011: Profile of Census Subdivisions (public-use microdata file)*. Ottawa ON: Statistics Canada. Using CHASS (distributor). <http://datacentre2.chass.utoronto.ca.cyber.usask.ca/cgi-bin/census/2011/>.
- Turgeon, S., Rodriguez, M.J., Thériault, M., Levallois, P., 2004. Perception of drinking water in the Quebec City region (Canada): the influence of water quality and consumer location in the distribution system. *Journal of Environmental Management* 70, 363–373. doi:10.1016/j.jenvman.2003.12.014



### **CHAPTER 3: RISK FACTORS ASSOCIATED WITH THE CHOICE TO DRINK BOTTLED WATER AND TAP WATER IN RURAL SASKATCHEWAN**

This manuscript was published in the International Journal of Environmental Research and Public Health. Copyright is held by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>). My contribution to the research included designing the sampling strategy, data collection and management, analysis, and writing the manuscript.

McLeod, L., Bharadwaj, L., Waldner, C., 2014. Risk Factors Associated with the Choice to Drink Bottled Water and Tap Water in Rural Saskatchewan. International Journal of Environmental Research and Public Health 11, 1626–1646. doi:10.3390/ijerph110201626

*The questionnaire used to establish a baseline for understanding the perceptions of water quality and risk was also used to investigate the drinking water preferences in rural Saskatchewan residents. Drinking water choices among rural populations in Saskatchewan have not been previously studied. Perceptions of quality and risk established in the previous chapter were investigated for influences on the choices rural residents make about drinking bottled water, and tap water, and in-home treatment of tap water. Understanding the drinking water choices of rural Saskatchewan residents, and what factors influence them, is important for the development of public health education about drinking water safety in rural areas. Knowledge of the factors influencing drinking water choices is also important to inform exposure assessments related to drinking water consumption in rural Saskatchewan residents.*

### **3.1 Abstract**

A cross-sectional study investigated risk factors associated with choices to drink bottled water and tap water in rural Saskatchewan. Of 7,500 anonymous postal questionnaires mailed out, 2,065 responses were analyzed using generalized linear mixed models. Those who reported a water advisory ( $p < 0.001$ ) or living in the area for  $\leq 10$  years ( $p = 0.01$ ) were more likely to choose bottled water. Those who reported tap water was not safe to drink were more likely to choose bottled water, an effect greater for those who had no aesthetic complaints ( $p < 0.001$ ), while those with aesthetic complaints were more likely to choose bottled water if they believed the water was safe ( $p < 0.001$ ). Respondents who treated their water and did not use a community supply were more likely to choose bottled water ( $p < 0.001$ ), while those who did not treat their water were more likely to choose bottled water regardless of whether a community supply was used ( $p < 0.001$ ). A similar pattern of risk factors was associated with a decreased likelihood of consuming tap water daily; however, the use of a community water supply was not significant. Understanding the factors involved in drinking water choices could inform public health education efforts regarding water management in rural areas.

### **3.2 Introduction**

According to a recent Canadian survey (Statistics Canada, 2011a), 20% of Saskatchewan residents reported choosing bottled water as their primary source of drinking water. Choices around drinking water consumption are governed by a complex set of factors relating to sensory perception, risk perception, and economic, psychological and social factors, including media reports and marketing messages (Doria et al., 2009). Additionally, accessibility and cost of bottled water are important factors, especially in rural and remote areas (Doria, 2006). Several studies have investigated a variety of risk factors associated with aspects of drinking water choices in North America, but few have considered the drinking water choices made by rural residents and we are not aware of any that have exclusively investigated drinking water choices of residents in rural areas of Canada. Previous studies have examined the influence of risk factors on choosing to drink bottled water (Hu et al., 2011; McSpirit and Reid, 2011; Saylor et al., 2011; Merkel et al., 2012), the risk factors associated with choosing bottled water and using in-home

treatment of tap water (Jones et al., 2006a , 2007b), and risk factors associated with choosing tap water, filtered tap water, or bottled water (Dupont et al., 2010).

Perceptions of water quality and risk are important factors in the choice to drink bottled water (Doria et al., 2009; Dupont et al., 2010; Hu et al., 2011; Saylor et al., 2011). Aesthetic qualities of water, particularly taste and odor, also appear to be associated with the choice to drink bottled water (Levallois et al., 1999; Doria et al., 2009; Dupont et al., 2010; Saylor et al., 2011).

Choosing bottled water has also been associated with age (Jones et al., 2007b; Dupont et al., 2010; Hu et al., 2011), gender (Dupont et al., 2010; Hu et al., 2011), and income (Jones et al., 2007b; Dupont et al., 2010). Though not examined in many studies, the household's water source could play a role in the choice to drink bottled water (Pintar et al., 2009; Hu et al., 2011), and regional differences have also been found (Jones et al., 2007b; Hu et al., 2011).

Many Canadian studies of drinking water consumption patterns have taken place in urban settings where water quality is routinely monitored, but in rural areas, residents may use a range of tap water sources, including private supplies for which the owner has sole responsibility for monitoring. These supplies can come from surface or groundwater sources of variable quality, and they can be impacted by local land use activities (Corkal et al., 2004). We hypothesized that types of water sources, water quality, and risk perception could be important factors influencing drinking water choices in rural Saskatchewan. The goal of this study was to gain a better understanding of how water sources and water quality and risk perception might influence choices around drinking water in rural Saskatchewan. Our primary objectives were to examine risk factors associated with the choices to consume tap water and bottled water in rural Saskatchewan. We also examined the factors associated with the choice to treat tap water using equipment in the home.

### **3.3 Materials and Methods**

#### *3.3.1. Design*

An anonymous postal questionnaire was administered to 7,500 rural households in six geographic regions of Saskatchewan in the fall of 2011. The questionnaire was distributed through Canada Post's Unaddressed AdMail service, which provides delivery of bulk mail without specific addresses to houses and farms within a given postal code. Target postal codes

were selected by using Canada Post Householder Counts in conjunction with postal code geography files (DMTI Spatial, 2006). A geographic information system (ArcMAP, ESRI, Redlands, CA, USA) was used to calculate the smallest radius around a central point selected for each region which would include the centroids of enough postal codes to encompass 1,250 eligible households. To ensure that the questionnaire would be distributed primarily to rural households, postal codes that did not include any farms were excluded, and where postal codes contained more than 200 houses, the survey was sent only to farms within that postal code.

Questionnaires were sent to 1,250 households from between nine and 12 postal codes (median=10) in each of the six regions for a total of 60 postal codes. The resultant data included a multistage, hierarchical sample of respondents from households within postal codes selected from within each geographic region. As a result of this distribution process, the questionnaire was delivered to a sample of residents from 24% of the rural municipalities within Saskatchewan.

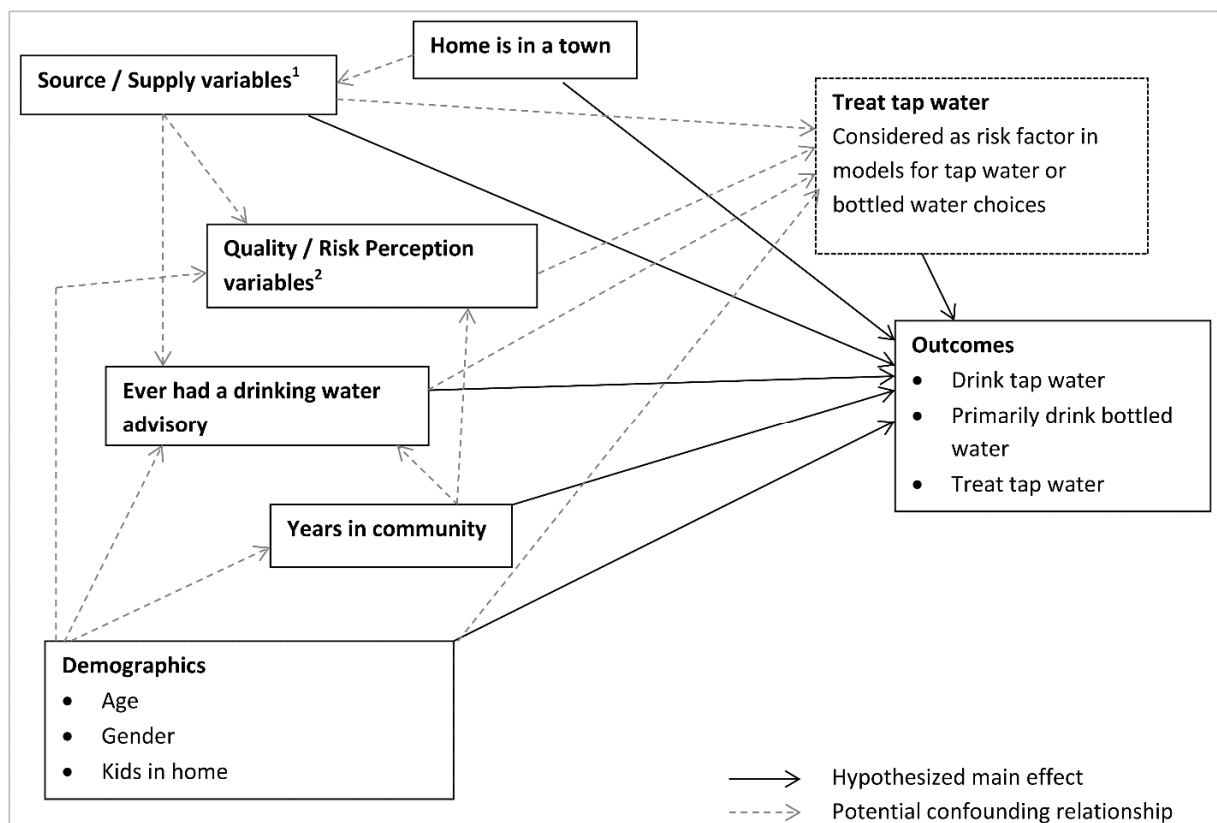
The four-page survey consisted of questions about household water sources, perceptions of quality and health risks from drinking water, consumption of tap water and bottled water, home treatment of tap water, and demographics. The questionnaire was modified from one used in a pilot study in 2010. We requested that the questionnaire be filled out by one member of the household who was over the age of 18, and returned in a pre-addressed, postage-paid envelope. Distribution of reminders or follow up to households that did not respond was not practical given that the questionnaires were not addressed to specific households. Ethics approval was obtained from the University of Saskatchewan Behavioral Research Ethics Board (Beh 10-184).

### *3.3.2 Outcomes and Potential Risk Factors of Interest*

The primary outcomes of interest were whether respondents reported primarily drinking bottled water and whether they consumed their household's tap water daily. Of secondary interest was the choice to treat the household tap water in some way.

A causal diagram (Figure 3.1) was constructed to help guide the process of model development. Primarily choosing bottled water was recorded as a dichotomous variable based on a question about consumption of purchased bottled water in the home with three possible responses. If respondents chose "yes, it is the primary drinking water source" they were classified as primarily

bottled water users, whereas those who chose “no” or “yes we drink it sometimes” they were classified as not using primarily bottled water. Daily consumption of tap water was also a dichotomous variable. Treating the tap water was evaluated as a dichotomous outcome based on the response to a question asking if the respondent had any equipment in their home to make the tap water better or safer to drink. Because the use of in-home treatment devices has been examined as a predictor of water consumption patterns in other studies (Jones et al., 2007b; Pintar et al., 2009; Roche et al., 2012), the use of water treatment in the home was also assessed as a risk factor in the models for bottled and tap water choices.



Notes: 1 Source/supply variables include use of a community (treated) water supply, private water supply, groundwater source, or surface water source (none of which are considered mutually exclusive); 2 Quality/risk perception variables include having any aesthetic complaint about the tap water, perception that water is not safe to drink, fear water will become contaminated, and perception that tap water has made anyone ill (similarly, none are mutually exclusive).

**Figure 3.1** Generalized causal diagram used to direct model development for each of the outcomes related to water consumption and treatment choices (potential interactions not diagrammed for simplicity).

Risk factors examined included variables related to household tap water sources: use of a community managed water supply, use of a private water supply, and whether the water source is

groundwater or surface water. In rural areas, households sometimes use more than one water source; therefore, community and private supplies were not mutually exclusive, nor were ground and surface water sources, so each of these variables was analyzed separately.

Risk factors related to water quality and risk were also evaluated, including reporting any aesthetic complaint, the perception that tap water was not safe to drink, the fear that the water supply will become contaminated, and the perception that the tap water had made anyone ill. Reporting any aesthetic complaint was a dichotomous variable, recoded from a question for which respondents could select any number of choices from a list of complaints about their tap water. If any of odor, bad taste, discoloration or cloudiness were selected, the respondent was considered as having any aesthetic complaint about their tap water. The perception that the water was not safe to drink, fear that water would become contaminated, and the perception that the tap water had made someone ill were dichotomous variables and were based on questions for which yes or no responses could be given.

Whether or not the household had ever experienced a drinking water advisory was analyzed as a dichotomous risk factor. Respondents reported whether or not an advisory had ever been experienced, but not reasons for the advisories or the time frame within which past advisories were experienced. The number of years residing in the current community, age, gender, and whether there are children in the household were also analyzed as risk factors. Six age categories were recorded on the questionnaire (18–24, 25–34, 35–44, 45–54, 55–64 and  $\geq 65$  years); however, due to low numbers of responses in the three youngest age groups, these were collapsed into a single category, so that only four age categories were used in the analysis (i.e., 18–44, 45–54, 55–64 and  $\geq 65$  years). Four possible categories for the number of years residing in the community (0–5, 6–10, 11–20 and 21 or more years) were also collapsed into two categories ( $\leq 10$  years,  $> 10$  years) for analysis.

### *3.3.3 Statistical Analysis*

Each of the outcomes was modeled using a generalized linear mixed model, specifying a binomial distribution and logit link function. Random intercepts were included in all models for both postal code and geographic region to account for any clustering arising from the hierarchical structure of the data.

Models were built for each outcome by first screening each risk factor individually where any risk factor with  $p < 0.2$  was retained for consideration when building the final model. Manual backwards selection was used to build the final main-effects model, retaining only risk factors with  $p < 0.05$ . All risk factors dropped from the main effects model were then assessed for confounding based on whether its inclusion in the model led to a change greater than 10% in the regression coefficients for other risk factors. Biologically plausible two-way interactions between risk factors retained in the final model were assessed at a 0.05 level of significance; in the case of categorical variables, a type 3 likelihood ratio test was used to determine if the interaction was significant.

Models were built in Stata (StataCorp LP, College Station, TX, USA) with the `xtmelogit` command using a Laplacian approximation for efficiency. Using the risk factors identified in the model building process, the final model parameters were estimated with `gllamm`, using adaptive quadrature with 12 integration points. Estimates of the random effects and predicted probabilities for each model were produced using `gllapred`. Population averaged probabilities were estimated using the `gllapred` marginal function (Rabe-Hesketh and Skrondal, 2008).

The proportion of the variance accounted for by postal code and region were examined for each of the outcomes, using an approximation of the variance partition coefficient for the binomial outcome based on the latent response variable model (Browne et al., 2005).

Values were missing for all outcomes and risk factors from at least one survey; any observations that were missing values for any of the risk factors or outcome for a given model were excluded from analyses including that variable. Therefore, the final number of observations used in each model varies and was reported for each model. Model assumptions were examined by evaluating the distribution of the residuals at the postal code and geographic region levels using Q-Q plots. The potential for outliers and influential data points was also investigated by plotting the standardized residuals at each level.

### **3.4 Results**

#### *3.4.1 Descriptive Statistics*

Of the 7,500 questionnaires sent out, 2,074 were returned. Seven were excluded because the postal code identifier had been removed by the respondent, one was excluded because it was blank, and one was excluded for being returned after the cut-off date for responses. As a result, 2,065 responses were used in the analyses, an effective response rate of 27.5%.

The median number of responses for each postal code was 44 (range 2–108) and the median number of responses per region was 353 (range 327–368). Frequencies were calculated for each outcome and risk factor (Table 3.1).



**Table 3.1** Number of complete and missing responses and the proportion of respondents at each level for the outcomes modeled and risk factors evaluated.

Variable	Complete	Missing	Response	Frequency	
	<i>n</i>	<i>n</i>		<i>n</i>	% <sup>1</sup>
<i>Outcomes</i>					
Primarily Drink Bottled water	2030	35	Yes	626	30.8
			No	1404	69.2
Drink Tap Water Daily	2013	52	Yes	1223	60.8
			No	790	39.2
Treat tap water in-home <sup>2</sup>	2003	62	Yes	953	47.6
			No	1050	52.4
<i>Risk Factors</i>					
Private water supply	2059	6	Yes	1249	60.7
			No	810	39.3
Community treated water supply	2059	6	Yes	640	31.1
			No	1419	68.9
Ground water source	1857	208	Yes	1349	72.6
			No	508	27.4
Surface water source	1856	209	Yes	613	33.0
			No	1243	67.0
Any aesthetic complaint about tap water	1984	81	Yes	501	25.3
			No	1483	74.8
Believe tap water not safe to drink	1984	81	Yes	235	11.8
			No	1749	88.2
Fear of contamination of water supply	1988	77	Yes	706	35.5
			No	1282	64.5
Anyone ever been ill from tap water	1784	281	Yes	57	3.2
			No	1727	96.8
Ever had water advisory	1981	84	Yes	485	24.5
			No	1496	75.5
Number of years in community	2046	19	≤ 10 years <sup>3</sup>	403	19.7
			> 10 years	1643	80.3
Home is in a town	2047	18	Yes	525	25.7
			No	1522	74.4
Gender	2005	60	Female <sup>3</sup>	1053	52.5
			Male	952	47.5
Age	2050	15	18-45 years <sup>3</sup>	317	15.5
			45-54 years	446	21.8
			55-64 years	614	30.0
			≥ 65 years	673	32.8
Children in the home	1932	133	Yes	437	22.6
			No	1495	77.4

<sup>1</sup>Proportion of complete observations. <sup>2</sup>Used as a risk factor in models for choosing tap water and choosing bottled water. <sup>3</sup>Reference category

To assess the representativeness of our sample, the distribution of age categories and gender in our sample were compared to data from the Canada 2011 Census of Population (Statistics Canada, 2011b) for the Census Subdivisions corresponding to the rural municipalities included in our survey regions (Table 3.2).

**Table 3.2** Comparison of frequency of key demographic variables in the survey sample population and the Statistics Canada 2011 Census of Population for the rural Census Subdivisions included within the survey regions.

Category	Survey respondents <sup>1</sup>		2011 Census of Population (%)
	N	(%)	
Female	1053	52.5	47.3
Male	952	47.5	52.7
18-45 years	317	15.5	35.7
45-54 years	446	21.8	23.2
55-65 years	614	30.0	21.7
≥65 years	673	32.8	19.3

<sup>1</sup>Total number of respondents: 2005 for gender and 2050 for age

#### 3.4.2 Choosing Primarily Bottled Water

With respect to drinking water preferences, 30.8% of respondents reported primarily consuming bottled water (Table 3.1). Of the respondents with a private water supply, 30.7% (376/1,224) reported primarily choosing bottled water, while 28.1% (178/634) of those using a community water supply reported consuming primarily bottled water. Use of other types of water supply were less common; 32.5% (39/120) of those who used a public water station (i.e., a community-maintained, publically available fill station) and 39.5% (47/119) of those whose water was delivered by truck reported primarily using bottled water.

After accounting for other significant risk factors, reporting a water advisory increased the likelihood of choosing primarily bottled water (OR=1.7,  $p<0.001$ ) compared to not reporting an advisory (Table 3.3, Figure 3.2). Of the respondents who reported ever having an advisory, 16.3% (79/485) of the respondents also reported having a current water advisory for their household. Of those that reported current advisories, 60.0% (42/70) reported drinking primarily bottled water and 34.8% (24/69) reported drinking their tap water daily.

Those who agreed that their tap water was not safe to drink were more likely to consume primarily bottled water than those who did not agree, but the effect of concern about unsafe tap water was greater for those who did not report any aesthetic complaints about the tap water (OR=8.5,  $p<0.001$ ) than for those who did (OR=2.3,  $p=0.001$ ) (Table 3.3, Figure 3.2). Compared to not reporting an aesthetic complaint, reporting any aesthetic complaint about tap water increased the likelihood of primarily consuming bottled water 6.1 times, but only for those who believed the tap water was safe ( $p<0.001$ ) (Table 3.3, Figure 3.2).

Those who lived in an area for  $\leq 10$  years were more likely to consume primarily bottled water than those who had not lived there as long (OR=1.5,  $p=0.01$ ) (Table 3.3, Figure 3.3).

Respondents who treated their tap water in their home and didn't use a community water supply were more likely to primarily consume bottled water compared those who treated their tap water and used a community supply (OR=2.5,  $p<0.001$ ) (Table 3.3, Figure 3.3). Compared to those who treated their tap water, respondents that did not treat their tap water were more likely to primarily consume bottled water, but the extent of the increase was greater for those who used a community water supply (OR=4.6,  $p<0.001$ ) than for those who did not (OR=2.5,  $p<0.001$ ) (Table 3.3, Figure 3.3).

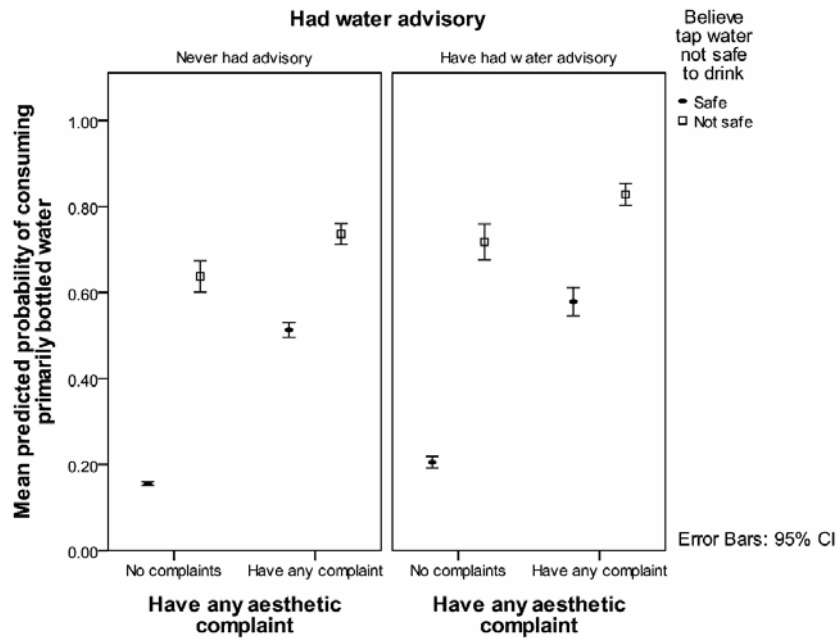
The proportion of variance explained by postal code (4.3%) was greater than that explained by region (0.3%) in the final multivariable model for primarily choosing bottled water. This represents a 40% improvement in the variance explained by postal code compared to the random effects of the null model, in which postal code accounted for 7.1% of the variance and region accounted for 0.4%.

**Table 3.3** Risk factors associated with the primary consumption of bottled water in the final multivariable model.

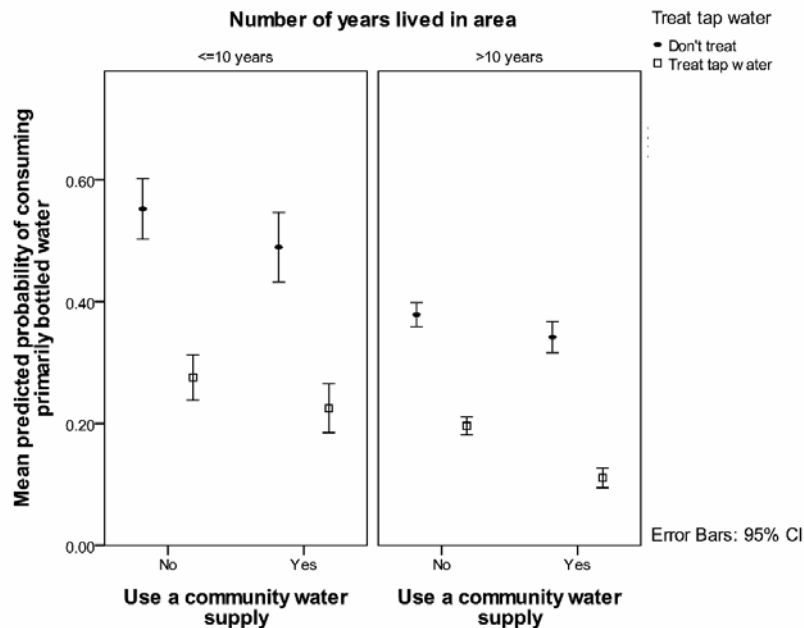
Risk factor	OR	95% CI		p
		lower	upper	
Ever had water advisory	1.7	1.3	2.4	< 0.001
Lived in area for > 10 years	Ref <sup>1</sup>			
Lived in area ≤ 10 years	1.5	1.1	2.0	0.01
Believe water not safe X have any aesthetic complaint				< 0.001 <sup>2</sup>
Believe tap water is not safe to drink, compared believing it is safe, for those who have any aesthetic complaint	2.3	1.4	3.8	0.001
Believe tap water is not safe to drink, compared to believing it is safe, for those who have no aesthetic complaints	8.5	5.2	13.9	< 0.001
Have any aesthetic complaint about tap water compared to not having a complaint, for those who believe their tap water is not safe	1.7	0.9	3.2	0.13
Have any aesthetic complaint about tap water compared to not having a complaint, for those who believe their tap water is safe	6.1	4.6	8.0	< 0.001
Use a community water supply X treat tap water				0.03 <sup>2</sup>
Not using a community water supply compared to using a community supply, for those who treat the tap water	2.5	1.5	4.0	< 0.001
Not using a community water supply compared to using a community supply, for those who do not treat their tap water	1.3	0.9	1.9	0.10
Not treating tap water compared to treating tap water, for those who use a community water supply	4.6	2.9	7.3	< 0.001
Not treating tap water compared to treating tap water, for those who do not use a community water supply	2.5	1.9	3.3	< 0.001
<b>Variances of Random Effects</b>				
	<b>Variance</b>	<b>SE</b>		
Postal code	0.147	0.086		
Region	0.010	0.030		

Number of observations=1844

<sup>1</sup>Reference category. <sup>2</sup>Overall p-value for interaction based on type 3 likelihood ratio test  
OR=odds ratio. CI=confidence interval. SE=standard error.



**Figure 3.2** Predicted probability of consuming primarily bottled water by presence of aesthetic complaint and the belief that tap water is not safe to drink, separated by whether or not household had a water advisory in the past averaged over all length of time residing in area, whether a community water supply is used, and in-home treatment of tap water.



**Figure 3.3** Predicted probability of consuming primarily bottled water by use of a community water supply and in-home treatment of tap water separated by length of residence in the area averaged over reported aesthetic complaints, agreement that tap water is not safe to drink and whether the household experienced a water advisory.

### *3.4.3 Consuming Tap Water Daily*

Most people (74.6%, (1,518/2,036)) reported consuming their tap water at least some of the time and 60.8% reported drinking tap water on a daily basis (Table 3.1). Of the respondents who used a private water supply, 63% (762/1216) reported consuming tap water daily. Daily tap water consumption was also reported by 61% (380/628) of respondents who used a community supply, 54% (63/117) of those who used truck-delivered water, and 55% (54/117) of those who used a public water station.

After accounting for other significant risk factors, reporting a water advisory decreased the likelihood of consuming tap water daily compared to not reporting an advisory (OR=0.7,  $p=0.004$ ) (Table 3.4). Those who lived in an area >10 years were 1.6 times more likely to consume tap water daily ( $p<0.001$ ) than those who lived in an area for a shorter time (Table 3.4).

**Table 3.4** Risk factors included in final multivariable model for daily consumption of tap water.

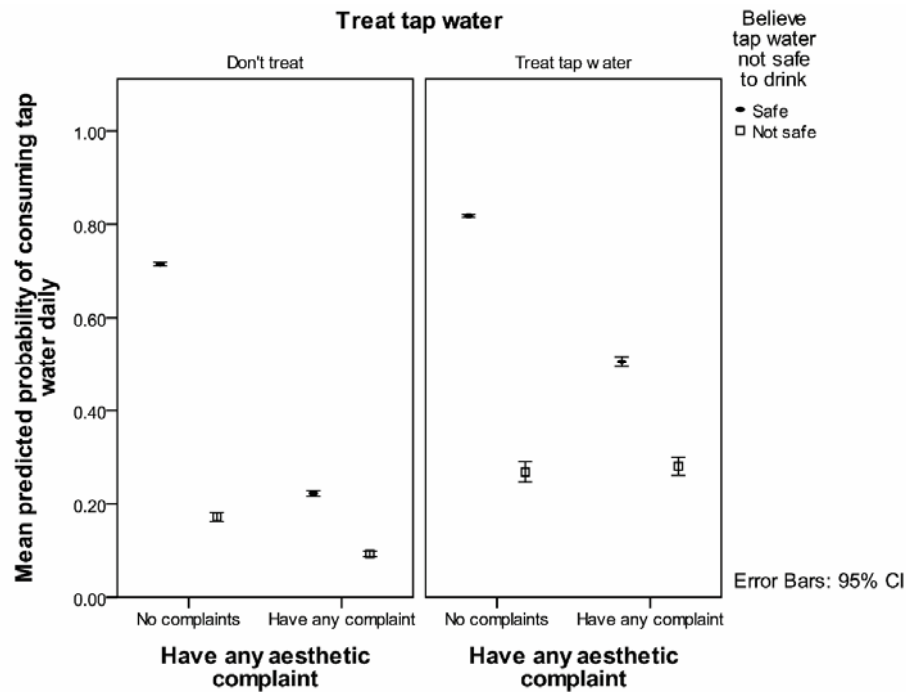
Risk Factor	OR	95% CI		p
		lower	upper	
Ever had water advisory	0.7	0.5	0.9	0.004
Lived in area <10 years	Ref <sup>1</sup>			
Lived in area > 10 years	1.6	1.2	2.1	0.001
Believe tap water not safe X any aesthetic complaint				0.001 <sup>2</sup>
Believe tap water not safe to drink compared to believing it is safe, for those with any aesthetic complaint	0.4	0.2	0.7	0.001
Believe that tap water not safe to drink compared to believing it is safe, for those with no aesthetic complaints	0.1	0.1	0.2	< 0.001
Have any aesthetic complaint compared to not having any aesthetic complaint, for those who believe the tap water is not safe	0.5	0.2	1.1	0.09
Have any aesthetic complaint compared to not having any aesthetic complaint, for those who believe the tap water is safe	0.1	0.1	0.2	< 0.001
Have any aesthetic complaint X treat tap water				0.006 <sup>2</sup>
Have any aesthetic complaint compared to not having any aesthetic complaint, for those who treat their tap water	0.2	0.2	0.3	< 0.001
Have any aesthetic complaint compared to not having any aesthetic complaint, for those who do not treat their tap water	0.1	0.1	0.2	< 0.001
Treat tap water compared to not treating tap water, for those with any aesthetic complaint	3.7	2.4	5.7	< 0.001
Treat tap water compared to not treating tap water, for those with no aesthetic complaints	1.8	1.4	2.4	< 0.001
<b>Variances of Random Effects</b>				
	<b>Variance</b>	<b>SE</b>		
Postal code	0.106	0.064		
Region	2.4x10 <sup>-15</sup>	6.4x10 <sup>-8</sup>		
Number of observations=1830				

<sup>1</sup>Reference category. <sup>2</sup>Overall p-value for interaction based on type 3 likelihood ratio test.

OR=odds ratio. CI=confidence interval. SE=standard error.

Those who did not think that their tap water was safe to drink were less likely to consume tap water daily than those who did, but the magnitude of the effect was slightly smaller for those who also reported aesthetic complaints (OR=0.4, p<0.001) than for those who did not (OR=0.1,

$p<0.001$ ) (Table 3.4, Figure 3.4). When compared to those who believed their tap water was safe and had no aesthetic complaints, respondents who thought their tap water was safe, but had at least one aesthetic complaint were 10 times less likely to consume their tap water daily ( $p<0.001$ ) (Table 3.4, Figure 3.4).



**Figure 3.4** Predicted probability of consuming tap water daily by presence of aesthetic complaint and the belief that tap water is not safe to drink, separated by in-home treatment of tap water averaged over all whether household had experienced a water advisory and length of time resided in area.

Reporting at least one aesthetic complaint also decreased the likelihood of consuming tap water daily for all respondents, with the effect of an aesthetic effect being greater for those who did not treat their tap water ( $OR=0.1$ ,  $p<0.001$ ) than for those who did ( $OR=0.2$ ,  $p<0.001$ ) (Table 3.4, Figure 3.4). Those who treated their tap water in some way were more likely to consume tap water daily compared to those who did not, but the importance of treatment was greater for those who also reported an aesthetic complaint ( $OR=3.7$ ,  $p<0.001$ ) than for those who had no complaints ( $OR=1.8$ ,  $p<0.001$ ) (Table 3.4, Figure 3.4).



The proportion of the variance in the final multivariable model for consuming tap water daily explained by postal code was 3.1%, while a negligible proportion was explained by region of the province ( $<0.001\%$ ). This represents a slight improvement compared to the random effects in the null model, in which postal code accounted for 4.9% of the variance and region accounted for 0.4%. The main effects in this model explained nearly 100% of the already small variation between regions of the province, and 38% of the variation between postal codes.

#### *3.4.4 In-home Treatment of Tap Water*

Of the respondents, 47.6% reported that they treat their household tap water in some way (Table 3.1). Of the respondents using a private water supply, 52.4% (637/1,215) reported treating their tap water, while 42.5% (264/621) of those who used a community water source reported treating their tap water in the home. Of those who used a surface water source, 41.6% (248/596) treated their water, while 50.7% (666/1,314) of those who used a groundwater source reported treating their tap water. Of the respondents who indicated which type of treatment they used, 58.7% (501/853) used a water softener, including 68.2% (416/610) of those using a groundwater source and 34.1% (76/233) of those using a surface water source. Of all respondents using water softeners, 72.1% (361/501) also indicated that some other form of water treatment (e.g., reverse osmosis, jug filter, ultraviolet, and distillation) was used.

After accounting for other risk factors, those who used a private supply were more likely ( $OR=2.1$ ,  $p<0.001$ ) to treat their household water compared to those who did not use a private supply (Table 3.5, Figure 3.5). Those who reported their tap water was not safe were half as likely to treat their tap water ( $p<0.001$ ) (Table 3.5, Figure 3.5) than those who did not. Having children under 18 residing in the home increased the likelihood of treating the tap water ( $OR=1.6$ ,  $p<0.001$ ) compared to not having children in the home (Table 3.5, Figure 3.5). Estimates were adjusted to minimize potential confounding by whether the home was in a town, and whether or not a community water supply was used.

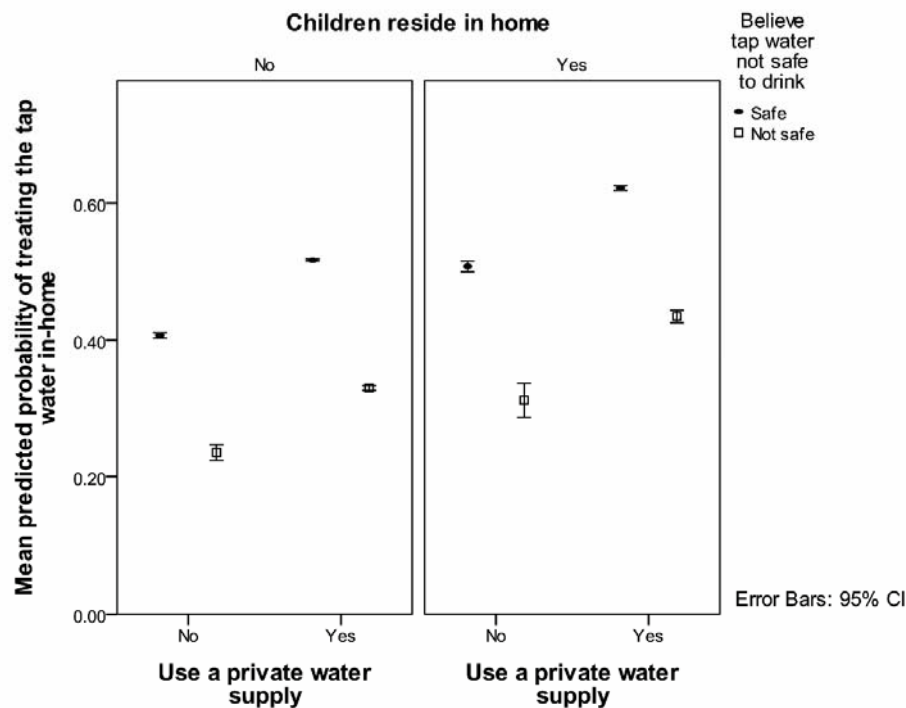
The random effects explained only a small proportion in the variation between postal code (2.0%) and region (1.1%) in the final multivariable model with little change from the proportions in the null model (2% and 1.5% respectively). This suggested there were few differences among postal codes and regions regarding the decision to treat water.

**Table 3.5** Risk factors included in final multivariable model for in-home treatment of tap water.

Risk Factor	OR	95% CI		p
Use a private water supply	2.1	1.5	3.0	< 0.001
Believe tap water not safe	0.5	0.3	0.6	< 0.001
Children reside in home	1.6	1.2	1.9	< 0.001
Home is in a town	1.2	0.8	1.7	0.31
Use a community water supply	1.3	0.9	1.8	0.25
Variances of Random Effects	Variance	SE		
Postal code	0.067	0.044		
Region	0.038	0.035		
Number of observations= 1796				

<sup>1</sup>Reference Category

OR=odds ratio. CI=confidence interval. SE=standard error.



**Figure 3.5** Predicted probability of treating tap water in-home by use of a private water supply and the belief that tap water is not safe to drink, separated by whether or not children reside in household, adjusted for whether the home was in a town and whether a community water supply was used.

The risk factors identified for consuming primarily bottled water, consuming tap water daily and choosing to treat tap water were summarized in Table 3.6.

**Table 3.6** Summary of the analyses for consuming primarily bottled water, consuming tap water daily and choosing to treat tap water, including the number of observations used and the risk factors associated with the final model for each outcome (effect modifications italicized).

<b>Outcome</b>	<b>n</b>	<b>Factors that increase likelihood of outcome</b>	<b>Factors that decrease likelihood of outcome</b>
Primarily choose bottled water	1711	History of water advisory <i>Believe tap water is not safe to drink – magnitude of effect greater for those that have no aesthetic complaints</i> <i>Have any aesthetic complaints about tap water, only for those who believe their tap water is safe</i>	Lived in area longer than 10 years <i>Use a community water supply, only for those who treat tap water</i> <i>Treat the tap water, with the magnitude of effect larger for those who use a community water supply</i>
Consume tap water daily	1818	Lived in area longer than 10 years <i>Treat tap water – magnitude of effect larger for those that also have aesthetic complaint about tap water</i>	History of water advisory <i>Believe tap water not safe to drink – magnitude of effect greater for those with no aesthetic complaint</i> <i>Have any aesthetic complaint, only for those who belief the tap water is safe to drink</i> <i>Have any aesthetic complaint - magnitude greater if tap water not treated</i>
Treat tap water <sup>1</sup>	1796	Use a private water supply Children reside in the home	Believe tap water not safe to drink

<sup>1</sup>Adjusted for whether home is in a town and whether a community water supply is used.

### 3.5 Discussion

In the present study, we investigated risk factors associated with water consumption choices among rural Saskatchewan residents, using quantitative analyses to examine the combined influence of several risk factors and account for the potential for clustering by community. The outcomes under investigation were not mutually exclusive, and the risk factors underlying the choice to primarily consume bottled water and to consume tap water daily were similar, although as expected their effects were opposite for these two outcomes. The models for primarily choosing bottled water and for daily consumption of tap water both included length of residence in the area, having had a water advisory, and an interaction between the belief that the tap water is not safe to drink and having any aesthetic complaint about the tap water. The risk factors for in-home treatment of the tap water included the use of a private water source, belief the water was not safe to drink, and whether children resided in the home.

#### *3.5.1 Consuming Primarily Bottled Water*

Just over 30% of our respondents reported using primarily bottled water, with little difference between those who used a community water supply and those who used a private water supply. A recent national Canadian survey (Statistics Canada, 2011a) reported that 20% of all Saskatchewan respondents reported consuming primarily bottled water, and that 19% of respondents using a municipal water supply and 39% of respondents using non-municipal supplies reported using primarily bottled water. However, 93% of the Saskatchewan respondents to the Statistics Canada survey (Statistics Canada, 2011a) reported using a municipal water supply, compared to just 31% of respondents in our rural study population.

We did not attempt to quantify the proportion of bottled water consumed that would define a respondent as choosing primarily bottled water; whereas, some water consumption studies have set a threshold of greater than 75% bottled water use (Jones et al., 2006a , 2007b). Other Canadian studies had rates of primary bottled water use ranging from 22% (Statistics Canada, 2011a) to 35% (Jones et al., 2006b). A recent national US study reported that just 13% of respondents reported using bottled water (Hu et al., 2011).

Among rural Saskatchewan residents, the choice to consume primarily bottled water at home appears to be mediated by a number of related factors. An interaction was identified between the

belief that the tap water is not safe to drink and whether the respondent had any aesthetic complaints about the tap water. Having any aesthetic complaint increased the likelihood of choosing primarily bottled water, but only for respondents who felt their tap water was safe. Those who felt their tap water was not safe to drink were consistently more likely to choose bottled water than those who felt it was safe regardless of aesthetic concerns.

Although this interaction between these risk factors is unique to the present study, our findings build on previous studies which reported that bottled water use was related to aesthetic complaints (Dupont et al., 2010; Jones et al., 2006b; Levallois et al., 1999; Saylor et al., 2011) and perception of health risks from tap water (Auslander and Langlois, 1993; Jones et al., 2006b; Dupont et al., 2010; Hu et al., 2011; McSpirit and Reid, 2011; Saylor et al., 2011; Merkel et al., 2012).

Another interaction was identified between use of a community water source and whether respondents use some sort of in home treatment for their tap water. People who used a community water supply compared to some other supply were less likely to choose bottled water, but this was only true for those who treated their tap water. Respondents that treated their tap water were also less likely to consume primarily bottled water than those who did not, but the difference was greater for those that used a community water supply. Home treatment was identified as a factor that reduces the likelihood of consuming primarily bottled water in previous studies (Jones et al., 2006a, 2007b). Its interaction with the use of a community water supply might reflect the use of home treatment devices to remove chlorine taste from tap water (Jones et al., 2007b).

Respondents who reported having a water advisory were more likely to primarily consume bottled water. To our knowledge this has not been previously investigated as a risk factor. Drinking water advisories in Saskatchewan are issued for a variety of reasons. For larger distribution systems (flow rate > 18,000 L/day) precautionary advisories were most commonly issued in the year prior to our survey for operational reasons such as depressurization of the system, line breaks, planned maintenance or high turbidity levels which could compromise treatment, while in smaller systems precautionary advisories more often resulted from positive bacteriological testing. In both types of systems, emergency boil water orders were most often

due to finding coliform bacteria in the water (Saskatchewan Ministry of Environment, 2011). It is possible that having experienced an advisory could introduce doubts about the safety of the household water. Because those respondents reporting an advisory included some respondents that were currently under an advisory, the possibility that being under a current advisory was driving the consumption of bottled water was considered. However, relatively few respondents (16%) who reported ever having a water advisory also reported a current water advisory for their household. Furthermore, of the respondents with a current advisory, over a third reported drinking their tap water daily. Although this suggests that respondents might drink their tap water despite being under an advisory, it is also possible that respondents reported their typical daily beverage choices as opposed to the choices made specifically during an advisory situation. Given the importance of experiences with water advisories in choices made around drinking water the impact of water advisories on drinking water habits should be studied in more depth.

Respondents that had lived in the area for longer than 10 years were less likely to choose bottled water. Familiarity with the tap water has been identified as an important factor in perception of water quality (Dietrich, 2006; Doria, 2010), and this result suggests that familiarity also reduces the likelihood that alternatives to tap water will be sought.

The role of demographic variables such as age and gender has been inconsistent in previous studies (Doria, 2006). Although age was unconditionally associated with primarily choosing bottled water in the present study, age was not included in our final model for primarily consuming bottled water. Some studies have reported that age as a significant predictor of bottled water usage (Jones et al., 2006a, 2007b; Hu et al., 2011) with the consumption of bottled water declining above age 30 in at least two studies. The population of respondents to our questionnaire was skewed toward older age groups which could have reduced the power of our study to detect differences between older and younger age groups. However, it is also possible that the effects of age are mediated through other risk factors included in our model.

The role of gender is less clear. In some previous analyses that examined the effects of gender in isolation, being female was associated with increased consumption of bottled water (Dupont et al., 2010; Saylor et al., 2011). However, using multivariable analysis, Dupont (2010) reported that males with children were more likely to be bottled water users than males without children,

while Hu (2011) reported that females had increased odds of drinking bottled water. However, gender was not a significant risk factor in another study (McSpirit and Reid, 2011) or in the present study. It is possible the effect of gender depends on other factors in the population under study or its effects are partly mediated through other risk factors such as perceptions of quality and risk.

### *3.5.2 Consuming Tap Water Daily*

More than 60% of our respondents reported drinking their tap water on a daily basis. We did not clarify if these respondents were primarily consuming tap water, but just 3% (40/1,213) of the respondents in this group also indicated that they primarily consume bottled water.

Consequently, daily consumption of tap water was the only measure we had available to classify respondents as regular users of tap water. Considering that some respondents may not typically consume water on a daily basis, we could have underestimated the number of respondents who primarily consume tap water. In a national Canadian survey (Statistics Canada, 2011a), 76% of Saskatchewan residents reported consuming primarily tap water. In the same survey, 78% of residents on a municipal supply reported drinking tap water, while only 49% of those with private water supplies chose primarily tap water. Among our respondents the proportion of respondents that reported drinking tap water daily was similar among users of private and community supplies.

The groups of risk factors included in the final model for choosing to drink tap water daily were similar, though not identical, to those for choosing primarily bottled water, but with opposite effects. Survey participants that reported water advisories were less likely to report daily tap water consumption, suggesting that this experience might reduce their confidence in the safety of tap water. Living in an area for longer than 10 years was perhaps a measure of familiarity with the water, and increased the likelihood of consuming tap water daily.

Reporting any aesthetic complaint decreased the likelihood of daily tap water consumption, but its effect was modified by both whether residents believed the tap water was safe, and whether residents treated their tap water; these interactions appeared to be independent of each other.

Reporting an aesthetic complaint made it less likely that tap water would be consumed regularly, but only when the tap water was considered safe. The belief that the tap water was not safe to

drink made it less likely that tap water would be chosen regardless of the presence of any aesthetic complaints. This interaction was the inverse of a similar interaction found in the model for choosing to consume primarily bottled water.

The effect of reporting an aesthetic complaint on the likelihood of consuming the tap water was somewhat mitigated by treating the tap water, and treating the water had a greater effect on the likelihood of consuming tap water when an aesthetic complaint was reported. This interaction was important to the likelihood of choosing to consume tap water daily; whereas, having an aesthetic complaint was not an important risk factor for the decision to treat the water on its own. It appears that aesthetic qualities are important to the decision to treat only in to the context of whether the tap water is consumed regularly.

The choice to treat tap water was evaluated as a risk factor for the consumption of tap water even though the direction of the causal relationship between treating tap water and drinking tap water is not clear. For instance, the tap water might be chosen because of the perception that treatment has made it more safe or palatable, or the decision to treat might be made if tap water is the only viable option for drinking water and it is perceived to not be safe or palatable unless treated.

Few previous studies have examined the risk factors associated with primarily choosing to drink tap water. Dupont et al. (2010) used analysis of variance to examine factors associated with the proportion of tap water consumed relative to filtered and bottled water in Canada, and found that the degree of concern about health risks from tap water was inversely related to the proportion of tap water consumed, as was the presence of various aesthetic concerns. This was similar to our results and underscores the importance of perception of quality and risk in making choices about drinking water.

Given the similarity between the models for primarily consuming bottled water and regularly consuming tap water, it might be reasonable to assume that similar factors, acting in opposite directions, play a role in each choice. However, there were some differences in the risk factors for each choice, and it has been hypothesized that choosing bottled water is not necessarily an alternative to choosing tap water, but may instead be considered an alternative to other pre-packaged beverages such as soda and juice (Doria, 2006; Jones et al., 2007a). The relative



importance of bottled water as an alternative to tap water compared to other beverages requires further investigation.

### *3.5.3 Treating Tap Water*

We also investigated the risk factors associated with the decision to treat the household tap water. Use of in-home water treatment devices has become common (Health Canada, 2003). Statistics Canada (2011a) reported that 50% of Saskatchewan residents indicated that they treat their tap water with a purifier, filter, or by boiling prior to consumption. More than 47% of our respondents indicated that they used any type of equipment in the home intended to make the tap water “better or safer to drink.” This number includes respondents that used water softeners, which are not recommended for the treatment of drinking water (Health Canada, 2003). In other Canadian studies, water softeners were included as treatment devices, and rates of water treatment were similar (Jones et al., 2006a, 2007b). In the present study, of those who indicated the type of treatment device used, 72% of respondents who used a water softener also used another device intended to treat drinking water.

Our results indicated that believing the tap water was not safe reduced the likelihood of treating the tap water. This is contradictory to another cross-Canada study that found health concerns increased the likelihood of consuming filtered tap water (Dupont et al., 2010). However, there could be a substantial difference between having general health concerns about water and the belief that the water is not safe. It is possible that if respondents felt their tap water was unsafe, they had no intention of consuming the water so did not treat it, or did not trust that home treatment devices would make their water safe.

Use of a private water supply increased the likelihood that water would be treated, a finding opposite to a study in British Columbia that reported fewer private source users than expected treated their tap water in an unconditional analysis (Jones et al., 2007b). This discrepancy may be related to the high rate of use of water softeners among our private water supply users. A study in Nova Scotia also reported that respondents with private water supplies were less likely to treat their water than those connected to a municipal supply (Janmaat, 2007). However, in that study the use of bottled water was considered a type of water treatment making it difficult to directly

compare their results with the present study, where bottled water consumption was considered separate from treatment.

We did not find an association between perception of poor aesthetic quality and the decision to treat the tap water among our respondents. This finding contradicted the study performed in Nova Scotia which found that treating household water was associated with the perception of lower water quality (Janmaat, 2007). However, as previously mentioned, the risk factors identified in the Nova Scotia study may differ from ours because we evaluated bottle water consumption separately from other types of treatment.

The presence of children in the home increased the likelihood that the tap water would be treated. Dupont et al. (2010) found a similar relationship but only for males, whereas gender was not included in our final model. Our model did include confounding variables, suggesting that the factors leading to water treatment are complex and deserving of further study, especially with respect to clarifying the factors related to treatment intended to make drinking water safer or more palatable compared to addressing the mineral content of the water.

#### *3.5.4 Limitations*

As previously discussed, our models for tap water and treatment were limited by self-reported measures of relative tap water consumption and the goals of treatment. It would have been ideal to be able to model the risk factors for choosing primarily tap water for comparison to choosing primarily bottled water, rather than comparison to drinking their tap water on a daily basis.

Previous studies have suggested that water consumption decreases with age (Jones et al., 2006a, 2007b; Roche et al., 2012), which could make daily tap water consumption an especially poor proxy for choosing primarily tap water in older age groups. Overall, the purposive nature of our regional sampling and a relatively low response rate, especially among younger age groups (Table 2), might limit the generalizability of our findings.

### **3.6 Conclusions**

By surveying residents of rural Saskatchewan in different communities and different regions, we were able to estimate the importance of some factors involved in drinking water choices among respondents who have access to a variety of water supplies. While our study provides information about the relationships between factors related to water supply and water quality and

risk perception and bottled water use in rural Saskatchewan, there are likely many other factors that are involved including accessibility, convenience, marketing, social cues, and concerns about environmental waste (Doria, 2006). We examined risk factors associated with the decision to regularly consume tap water. While these were similar to those involved in influencing the choice to drink bottled water, it has also been suggested that consumers don't necessarily view bottled water as an alternative to tap water, but to other types of beverages such as soda and juice (Doria, 2006; Jones et al., 2007a). Further investigation of specific perceptions related to water quality and risk, especially in conjunction with estimates of the relative amounts of bottled water, tap water, and other beverages consumed is needed to better understand the drinking water and beverage choices made by residents of rural Saskatchewan. A better understanding of the factors involved in such decisions, and any regional differences in these factors, are crucial for informing public health efforts regarding the safety, testing and treatment of drinking water, as well as the assessment of health risks related to water consumption in rural areas.

### 3.7 References

- Auslander, B.A., Langlois, P.H., 1993. Toronto tap water: perception of its quality and use of alternatives. *Can. J. Public Health Rev. Can. Santé Publique* 84, 99–102.
- Browne, W.J., Subramanian, S.V., Jones, K., Goldstein, H., 2005. Variance partitioning in multilevel logistic models that exhibit overdispersion. *J. R. Stat. Soc. Ser. A Stat. Soc.* 168, 599–613.
- Corkal, D., Schutzman, W.C., Hilliard, C.R., 2004. Rural water safety from the source to the on-farm tap. *J. Toxicol. Environ. Health A* 67, 1619–1642.
- Dietrich, A.M., 2006. Aesthetic issues for drinking water. *J. Water Health* 4, 11–16.
- DMTI Spatial, 2006. DMTI Platinum Postal Suite version 2006.4: Saskatchewan Local Delivery Unit Area (LDU) boundaries. DMTI Spatial, Markham, ON.
- Doria, M. de F., 2010. Factors influencing public perception of drinking water quality. *Water Policy* 12, 1–19. doi:10.2166/wp.2009.051
- Doria, M. de F., Pidgeon, N., Hunter, P.R., 2009. Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study. *Sci. Total Environ.* 407, 5455–5464.
- Doria, M. de F., 2006. Bottled water versus tap water: understanding consumers' preferences. *J. Water Health* 4, 271–276.
- Dupont, D., Adamowicz, W.L., Krupnick, A., 2010. Differences in water consumption choices in Canada: the role of socio-demographics, experiences, and perceptions of health risks. *J. Water Health* 8, 671–686.
- Health Canada, 2003. Water Treatment Devices for the Removal of Taste, Odour and Chemicals. URL <http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/devices-dispositifs-eng.php>.
- Hu, Z., Morton, L.W., Mahler, R.L., 2011. Bottled water: United States consumers and their perceptions of water quality. *Int. J. Environ. Res. Public. Health* 8, 565–578.
- Janmaat, J., 2007. A Little Knowledge...: Household Water Quality Investment in the Annapolis Valley. *Can. J. Agric. Econ. Can. Agroéconomie* 55, 233–253. doi:10.1111/j.1744-7976.2007.00090.x
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., 2006a. Drinking water consumption patterns of residents in a Canadian community. *J. Water Health* 4, 125–138.
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Henson, S.J., Mathews, E., 2007a. A qualitative exploration of the public perception of municipal drinking water. *Water Policy* 9, 425–438. doi:10.2166/wp.2007.019
- Jones, A.Q., Dewey, C.E., Doré, K., Majowicz, S.E., McEwen, S.A., Waltner-Toews, D., Mathews, E., Carr, D.J., Henson, S.J., 2006b. Public perceptions of drinking water: a postal survey of residents with private water supplies. *BMC Public Health* 6, 94–104. doi:10.1186/1471-2458-6-94
- Jones, A.Q., Majowicz, S.E., Edge, V.L., Thomas, M.K., MacDougall, L., Fyfe, M., Atashband, S., Kovacs, S.J., 2007b. Drinking water consumption patterns in British Columbia: An

- investigation of associations with demographic factors and acute gastrointestinal illness. *Sci. Total Environ.* 388, 54–65. doi:10.1016/j.scitotenv.2007.08.028
- Levallois, P., Grondin, J., Gingras, S., 1999. Evaluation of consumer attitudes on taste and tap water alternatives in Quebec. *Water Sci. Technol.* 40, 135–139.
- McSpirit, S., Reid, C., 2011. Residents' Perceptions of Tap Water and Decisions to Purchase Bottled Water: A Survey Analysis from the Appalachian, Big Sandy Coal Mining Region of West Virginia. *Soc. Nat. Resour.* 24, 511–520.
- Merkel, L., Bicking, C., Sekhar, D., 2012. Parents' Perceptions of Water Safety and Quality. *J. Community Health* 37, 195–201. doi:10.1007/s10900-011-9436-9
- Pintar, K.D.M., Waltner-Toews, D., Charron, D., Pollari, F., Fazil, A., McEwen, S.A., Nesbitt, A., Majowicz, S., 2009. Water consumption habits of a south-western Ontario community. *J. Water Health* 7, 276–292.
- Rabe-Hesketh, S., Skrondal, A., 2008. *Multilevel and Longitudinal Modelling Using Stata*. Stata Press.
- Roche, S.M., Jones, A.Q., Majowicz, S.E., McEwen, S.A., Pintar, K.D., 2012. Drinking water consumption patterns in Canadian communities (2001-2007). *J. Water Health* 10, 69–86.
- Saskatchewan Ministry of Environment, 2011. 2010-11 State of Drinking Water Quality in Saskatchewan.
- Saylor, A., Prokopy, L.S., Amberg, S., 2011. What's wrong with the tap? Examining perceptions of tap water and bottled water at Purdue University. *Environ. Manage.* 48, 588–601.
- Statistics Canada, 2011a. *Households and the Environment: Analysis*. URL <http://www.statcan.gc.ca/pub/11-526-x/2013001/part-partie1-eng.htm>.
- Statistics Canada, 2011b. *Census of Canada 2011: Profile of Census Subdivisions (Public-use Microdata File)*. Ottawa ON: Statistics Canada. Using CHASS (distributor). <http://datacentre2.chass.utoronto.ca.cyber.usask.ca/cgi-bin/census/2011/>.

## **CHAPTER 4: EVALUATION OF THE USE OF EMPIRICAL BAYESIAN KRIGING FOR SUMMARIZING GROUNDWATER ARSENIC CONCENTRATIONS AND THE USE PRINCIPAL COMPONENTS ANALYSIS COMBINED WITH EMPIRICAL BAYESIAN KRIGING TO SUMMARIZE MIXTURES OF TRACE METALS AND MAJOR IONS IN GROUNDWATER IN SOUTHERN SASKATCHEWAN**

*The research outlined in this chapter summarizes observed water quality from groundwater supplies in Saskatchewan. Having established a baseline understanding of water perceptions, it was important to investigate the measurable quality of water supplies experienced by residents of rural Saskatchewan. Existing data collected as part of water quality surveillance activities was identified as a rich source of data to investigate drinking water quality in rural Saskatchewan. Data from public water supplies and private wells was examined using principal components analysis to summarize groups of contaminants identified in the Saskatchewan Drinking Water Quality Standards and Objectives. This was followed up by an investigation of the use of common geostatistical methods to interpolate groundwater quality across a large region of Saskatchewan. The innovative methods used here to summarize water quality are useful for informing policy designed to improve water quality for those who access groundwater supplies. In addition, the results were subsequently used to estimate exposures in epidemiological studies of associations between water quality and important chronic diseases.*

#### 4.1 Abstract

Groundwater drinking supply surveillance data were accessed to summarize water quality in public and private groundwater drinking supplies in southern Saskatchewan. The public water supply data were collected through routine monitoring of regulated public water supplies. De-identified private well data were accessed from a government-run program which offered well water management advice and subsidized water testing to landowners accessing unregulated, private water supplies. Arsenic in drinking water has been linked to a variety of chronic diseases; therefore, arsenic concentrations were analyzed in isolation. Other contaminants (e.g., heavy metals, major ions) were grouped according to their classification in the Saskatchewan Drinking Water Quality Standards and Objectives. Standards refer to contaminants which are considered potential health risks, while objectives relate to the aesthetic qualities of drinking water. Principal components analysis was applied to obtain principal component (PC) scores to summarize mixtures of correlated parameters identified as health standards and those identified as aesthetic objectives. The data were divided into training and validation data sets. Ordinary, universal, and empirical Bayesian kriging were used to interpolate arsenic concentrations and PC scores in southern Saskatchewan using the training data. Agreement between arsenic concentrations and PC scores predicted by kriging and the observed values in the validation data set was assessed using concordance correlation coefficient (CCC). In addition, agreement at predicting values above the 75<sup>th</sup> percentile in the observed data was assessed using kappa and sensitivity. For public water supplies, values of CCC ranged from 0.01 to 0.35 and from -0.001 to 0.52 for the private well data. Values of kappa ranged from -0.08 to 0.19 for the public water supply data and 0.0 to 0.39 for the private well data. Sensitivity for predicting values above the 75<sup>th</sup> percentile ranged from 0.0 to 0.33 for the public water supply data, and 0.0 to 0.38 for the private well data. While the measures of agreement between predicted and observed arsenic concentrations and PC scores were generally low, empirical Bayesian kriging performed best overall, based on having the greatest number of highest values of CCC, kappa, and sensitivity across all variables. All of the kriging methods appeared to underestimate high values of arsenic and PC scores. While not efficient at predicting high values, empirical Bayesian kriging is a valuable tool for summarizing large scale geographic trends in groundwater quality.

## 4.2 Introduction

Recent studies have highlighted associations between exposure to arsenic from drinking water and a variety of chronic diseases including diabetes mellitus, hypertension, cardiovascular disease and cancer (Navas-Acien et al., 2005; Chen et al., 2007; Celik et al., 2008; Lisabeth et al., 2010; Chen et al., 2011; Moon et al., 2012). It has also been suggested that poor water quality, especially aesthetic issues that impact consumer acceptance of tap water, might have indirect impacts on the development of chronic disease by motivating the consumption of sugar-sweetened beverages as an alternative to drinking water (Onufrak et al., 2014). Access to affordable tap water alternatives such as bottled water may be limited in rural and remote areas as compared to that available in urban locations, exacerbating the potential impacts of tap water with poor aesthetic quality for those located outside urban centres.

In Canada, drinking water oversight is a provincial mandate. Guidelines for acceptable levels of contaminants, as well as regulations pertaining to water quality testing, are established by each province but typically follow recommendations set forth in the federal Guidelines for Canadian Drinking Water Quality (Dunn et al., 2014; Health Canada, 2014). In the province of Saskatchewan (SK) there are three tiers of drinking water systems which are subject to different levels of regulation. Public drinking water systems are defined as those that feed a distribution system and have a design flow of greater than 18 cubic meters per day; these are monitored and regulated by the Saskatchewan Water Security Agency. Semi-public supplies are those that serve members of the public but lack a distribution network or are connected to at least three but fewer than 15 service connections and have a flow rate of less than 18 cubic meters of water per day; these systems are monitored by the Ministry of Health through local Health Regions. Private water supplies serve a single household or farm; these supplies are not regulated and have no monitoring requirements. Residents who use private water supplies bear sole responsibility for ensuring the safety and quality of their household water supply; however, they may lack the knowledge and resources to routinely and adequately test their household water source. As a consequence of inadequate testing of private wells, residents of rural areas could be at greater risk of exposure to drinking water of poor quality or that has elevated concentrations of toxins such as arsenic. For public supplies, the requirements for testing, including the list of parameters and sampling frequency, are based on the size of the population served (Water Security Agency, 2016). Smaller water distribution systems may suffer from a lack of resources and highly trained



personnel when compared to water treatment facilities in urban centers. Smaller systems may also lack the resources to avoid placing wells in aquifers of lower water quality. Therefore, residents in remote and rural areas, whether on a smaller public system or using a private water supply, may be vulnerable to health effects related to the quality of water available at their household tap.

Saskatchewan's Drinking Water Quality Standards and Objectives (Water Security Agency, n.d.) apply to regulated water supplies. Standards are legally enforceable and are based on parameters that are potentially harmful to human health. Standards refer to trace metals such as arsenic, boron, barium, cadmium, chromium, mercury, lead, selenium, and uranium, as well as a range of other chemical contaminants such as nitrate, cyanide, and pesticides. Standards also include microbiological and radiological contamination. Objectives are guidelines based on parameters that are not considered harmful to human health in the context of expected concentrations in drinking water, but which can negatively impact the aesthetic qualities of the water, such as odor, taste, and discoloration, that might impact the palatability and acceptance of drinking water to consumers. Objectives include ions such as sodium, potassium, chloride, sulphate, metals including iron and zinc, and other measures such as alkalinity, pH, total dissolved solids, and hardness.

Ground water quality has been previously identified as a public health concern in the province of Saskatchewan, Canada, particularly for residents using private wells for their drinking water (Thompson et al., 1999; Thompson, 2001; Thompson, 2003; Corkal et al., 2004; Charrois, 2010). It is estimated that approximately 28 percent of Saskatchewan residents use groundwater-based municipal or communal drinking water supplies, and another 15 percent of residents obtain drinking water from private water supplies (Government of Saskatchewan, 2009).

Groundwater, depending on local geology, is subject to contamination by leaching from minerals in geological formations as well as anthropogenic contamination by surface run off and infiltration through soil. A small study of arsenic levels in 61 private and rural municipality-owned wells in Saskatchewan found that 23 percent of the wells had arsenic concentrations above the current standard of 0.01 mg/L (Thompson et al., 1999). Additionally, at least one potential "hot spot" or cluster of wells with elevated arsenic concentrations was identified. Another study found that 99.6% of 535 wells sampled exceeded at least one aesthetic objective

or health standard; of those, 35% exceeded a health-related standard (Sketchell and Shaheen, 2000). In addition, other minerals and ions such as iron and manganese, though not considered health risks, are frequently found in groundwater at concentrations that exceed recommended levels and negatively affect the aesthetic quality of drinking water (Sketchell and Shaheen, 2000; Thompson, 2003). In a recent survey of residents in rural Saskatchewan 25 percent of respondents reported having an aesthetic complaint about their household drinking water (McLeod et al., 2015).

This study attempts to investigate and summarize groundwater composition in rural areas of Saskatchewan. The motivation for this work was to summarize groundwater quality in Saskatchewan for public and private water supplies in rural areas, and to identify patterns of groundwater composition throughout the province as a tool to inform public health recommendations for the testing of groundwater drinking water supplies. This is especially important for owners of private water systems who must prioritize testing for various parameters in the face of limited resources for testing. Additionally, results from this study will be used in an analysis of associations between water quality and rates of type 2 diabetes, hypertension, ischemic heart disease, and stroke in Saskatchewan. Estimating exposure to constituents of drinking water is a major challenge in investigating such associations.

Previous work has suggested that geostatistics can be used to map estimated exposure to arsenic through groundwater as a continuous surface across a region based on a limited set of point measures (Goovaerts et al., 2005; Meliker et al., 2008; Yang et al., 2009; Gong et al., 2014; James et al., 2014). Additionally, principal components analysis has been used as a variable reduction method to investigate groupings of groundwater parameters (Helena et al., 2000; Liu et al., 2003; Chapagain et al., 2010; Belkhiri et al., 2011) and has been combined with geostatistical methods to map underlying latent processes contributing to overall water quality across a region (Sánchez-Martos et al., 2001; Satyaji Rao et al., 2009; Shyu et al., 2011; Nazzal et al., 2015).

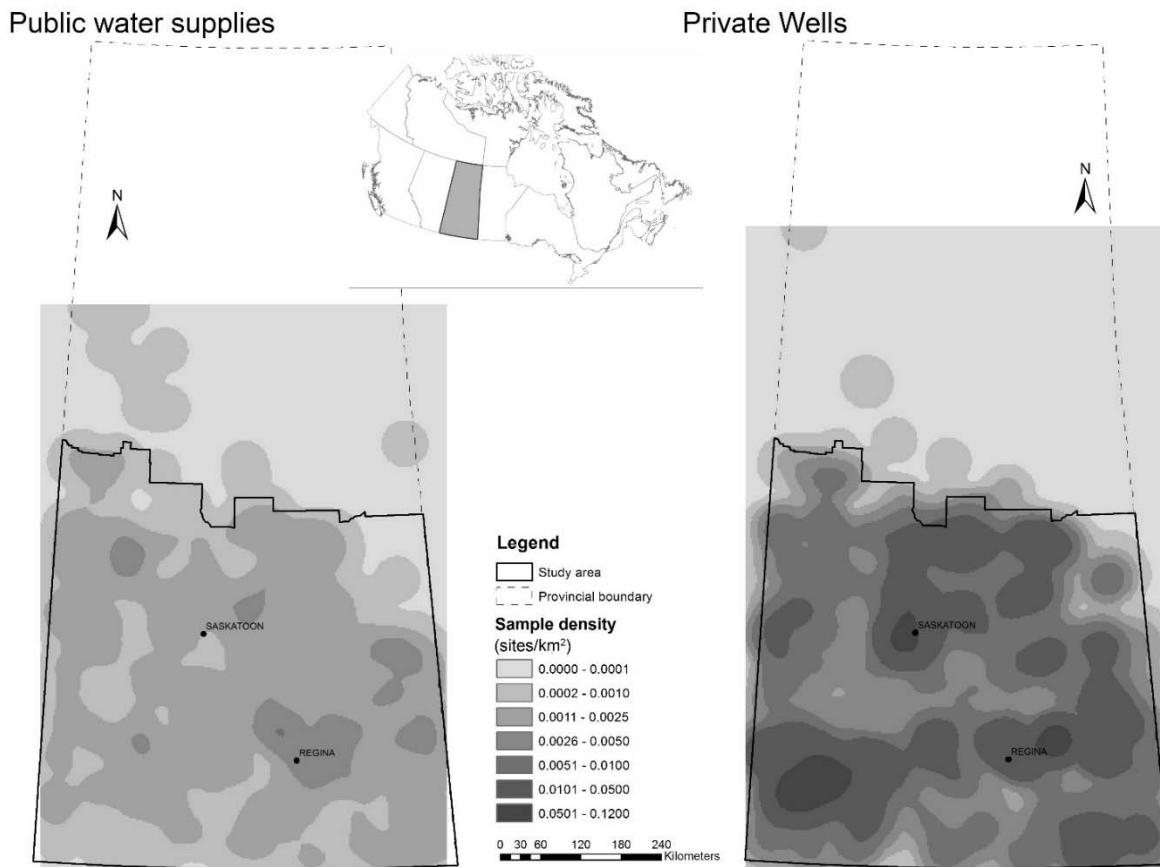
Previously collected water surveillance records represent an existing source of data by which water quality can be summarized and used to inform drinking water management recommendations. Our objective was to use water quality monitoring data from public systems and private water supplies to summarize water quality through principal components analysis of trace metals and major ions that are monitored as health standards or aesthetic objectives. Our

secondary objectives were to compare the performance of ordinary kriging, universal kriging, and empirical Bayesian kriging for predicting arsenic concentrations and principal component scores derived from principal components analysis, and to evaluate the potential use of these tools in human health exposure analysis for large rural areas of western Canada. The results of this analysis will be used to estimate drinking water arsenic concentrations and principal components summarizing water health standard and aesthetic parameters for geographic regions across rural Saskatchewan.

## **4.3 Materials and Methods**

### *4.3.1 Study Area*

Saskatchewan is a province located in central Canada with borders approximately located along 49° N and 60° N, and 102° and 110° W. The northern part of the province is sparsely populated, and because the water monitoring data were very sparse in the northern part of the province, the study area was limited to the southern part of Saskatchewan where sufficient data were available. The limits of the study area are bounded by the solid line in Figure 4.1. The irregular northern border of this area corresponds to the administrative boundaries of rural municipalities along this edge of the study area.



**Figure 4.1** The study area (solid line) within the province of Saskatchewan along with the density of sampling sites for public water supplies (left) and private wells (right) across the province. Inset map shows location of Saskatchewan within Canada.

#### 4.3.2 Data Sources

Analysis was carried out separately for public water supply data and private well data. All data were obtained from the Saskatchewan Water Security Agency. The public supply data originated from groundwater sourced systems that fall under regulation of the Water Security Agency and were collected from 1985-2012, and consisted primarily of treated water samples. The private water supply data were collected as part of the Water Security Agency's Rural Water Quality Advisory Program (RWQAP), a unique program designed to provide advice and subsidized water testing services to private well owners. All analyses were performed using methods accredited by the Canadian Association for Laboratory Accreditation Inc. (Canadian Association for Laboratory Accreditation Inc., n.d.) at the Saskatchewan Disease Control Laboratory

(Regina, SK) or the Saskatchewan Research Council Environmental Analytical Laboratories (Saskatoon, SK). The densities of sampling sites for public water supplies and private wells are shown in Figure 4.1.

Concentrations for all parameters, for both private and public supply databases, were converted to mg/L. Only samples from groundwater sources were retained, and samples with missing location information were discarded. Because analytic methods and detection limits changed over time, data were compared to the detection limits corresponding to the sampling date; for samples where the parameter concentration was reported as below the detection limit, half the detection limit was imputed for the missing value (Farnham et al., 2002; Meliker et al., 2008). The distributions for all parameters were right skewed and natural logarithm transformations were applied prior to analysis. Descriptive statistics for parameters of interest, along with the proportion of samples exceeding Saskatchewan's standards or objectives, and the proportion of samples that were below detection limits were summarized for the public water supply data and the private well data.

Cyanide and mercury were sampled less frequently than the other metals and ions of interest in public systems, while cyanide was not sampled in the private wells and mercury was sampled infrequently. Therefore, these parameters were excluded from the analysis for both public supplies and private wells to maintain consistency in the analysis of the different types of supplies. Additionally, 98 percent of cadmium samples and 72 percent of chromium samples were below detection limit for the public water supply data, and 92 percent of cadmium samples and 91 percent of chromium samples were below detection limit in the private well data. Consequently, cadmium and chromium were also excluded from the final analysis. Fluoride was infrequently reported for private wells, but was added to some municipal supplies (Dental Health Promotion Working Group of Saskatchewan, 2011) and was also excluded from the principal components analysis.

The water quality data for the public water supplies consisted of repeated measures taken at irregular intervals during the period 1985 to 2012. Because the kriging methods used in ArcGIS 10.2 (ArcGIS, ESRI, Redlands, CA) required a single measure at each point location, a linear mixed model was developed for each water parameter to estimate a single predicted value for each parameter for each water supply location. The model for each parameter included a random

intercept to account for between supply variation, as well as a structured error term to account for repeated sampling within water supplies.

The private well data consisted of measures at a single point in time for each location. However, due to privacy considerations, private well locations were generalized to the centroid of the section of land on which they were located before confidential data access was granted for this analysis. A section is approximately 1.6 km by 1.6 km and corresponds to a parcel of land described by the grid-based land description system used in western Canada (ISC, n.d.). Because these measures ultimately represented distinct wells at different locations, wells with duplicated generalized locations were manually jittered by alternately increasing or decreasing the latitude and longitude by increments of 10 m for each subsequent duplicate well location until no duplicate locations remained.

#### *4.3.3 Principal Components Analysis*

Principal components analysis (PCA) is a multivariate statistical technique that can be used to reduce the dimensionality of a data set with numerous correlated variables and categorize variables into groups based on their covariance. The resultant independent principal components retain all original information from the data except for that in any components that are not retained for subsequent analyses. PCA was performed using SAS (SAS Institute Inc., Cary, NC, USA) and applied separately to the public supply and private supply data, separated into groups of water parameters identified as health standards and as aesthetic objectives according to the Saskatchewan Drinking Water Quality Standards and Objectives. For the public water supply data, the single predicted value for each supply was used in PCA, while values for each individual well were used in the PCA of private well data.

Log transformed concentrations (mg/L) were converted to standardized z-scores prior to performing PCA. Kaiser's measure of sampling adequacy (KMO) and Bartlett's test of sphericity were performed to assess the sampling adequacy and correlation of these data and therefore their suitability for PCA. In each analysis, principal components (PC) with an eigenvalue of greater than one were retained, and subject to varimax rotation to maximize the variation explained by each of the retained principal components and to obtain the final principal component loadings and coefficients. Principal component (PC) scores for each of retained principal components were calculated for the public supplies and private wells for use in the geostatistical analysis.

#### *4.3.4 Geostatistical Analysis*

Kriging is a geostatistical method that incorporates spatial autocorrelation in continuous variables to interpolate values at locations at which they have not been measured. The basis of kriging is the semivariogram model, which uses the semivariance between point measures to summarize the spatial relationships in variables. Ordinary kriging and universal kriging are methods for which estimation is based on weighted least squares and the assumption that the calculated semivariogram is the single true semivariogram model for the data. Ordinary kriging differs from universal kriging in that ordinary kriging assumes a constant unknown mean across a given area, while universal kriging assumes a constant trend in the data.

Empirical Bayesian kriging first became available in ArcGIS version 10.1, and is based on restricted maximum likelihood estimation. Additionally, it allows for uncertainty in the semivariogram model by a process of data subsetting and simulation to estimate a range of semivariogram models. As a result, empirical Bayesian kriging requires fewer assumptions about the semivariogram model form and parameters, is considered accurate for moderately non-stationary data, and provides more accurate estimation of prediction standard errors than other kriging methods (Krivoruchko, 2012).

Principal component scores and log transformed arsenic concentrations were mapped in ArcGIS 10.2 (ArcGIS, ESRI, Redlands, CA) to compare the performance of ordinary, universal and Empirical Bayesian Kriging for the prediction of arsenic concentrations, PC scores for health standards, and PC scores for aesthetic objectives. The analyses for each of these three types of data was performed separately for public water supply data and private well data, making a total of 6 data sets to be evaluated with each kriging method. For development of kriging models, each dataset was randomly divided into training (70 percent of the observations) and validation (30 percent) datasets, such that a different subset of sites made up each of the six training datasets.

Variogram models for each variable in the training datasets were developed using SAS software (SAS Institute Inc., Cary, NC). First or second order large scale trends were identified in the data for each variable using regression analysis. Where trends were detected, the residuals from the regression models were used to develop and compare variogram models. The average nearest neighbor distance for the sampling locations for each variable were calculated in ArcGIS 10.2

and set as the lag distance for the variograms, and half the study area extent divided by the lag distance was used to calculate the maximum number of lags, which limits the semivariogram lag to half the extent of the sampling area (Olea, 2006). The fits of spherical, exponential, and Gaussian models were compared using the Akaike Information Criterion (AIC) for each variable in each of the public and private water supply data sets. The presence of anisotropy was evaluated visually by dividing the variogram into eight directions (22.5° angles). The exploratory spatial data analysis tools in ArcGIS were used to visually evaluate normality of the water data using quantile-quantile (Q-Q) plots and stationarity was assessed using Voronoi maps and semivariograms.

ArcGIS 10.2 was used to perform ordinary, universal and empirical Bayesian kriging for each variable using the training data sets. The order of trend removal, lag distance, and number of lags were set based on the variogram models developed in SAS, and the model form was specified as spherical for each variable for ordinary and universal kriging. Because empirical Bayesian kriging is less interactive and spherical models are not available, the settings for empirical Bayesian Kriging were left at default values, such that subsets of 100 points and 100 simulations were modeled for each variable using a power-based model.

The resultant prediction maps for each kriging technique were used to extract predicted values at the locations of the samples in the test data sets for each variable. Agreement between the predicted and actual values was assessed in Stata (StataCorp LP, College Station, TX) using Linn's concordance correlation coefficient (CCC) (Lin, 1989) and Spearman's rank correlation on the raw values, as well as the Kappa statistic on a dichotomized variable based on whether or not the predicted or actual value of the log arsenic concentration for each of the principal component scores were above the 75<sup>th</sup> percentile for the observed data. The dichotomized variable was also used to assess the sensitivity and specificity of each kriging model in predicting whether the log arsenic concentration and each of the principal components scores were above the 75<sup>th</sup> percentile for the observed data. For arsenic concentrations, the kappa statistic, sensitivity, and specificity were also calculated for whether the concentration was above the drinking water standard (0.01 mg/L). Limits of agreement plots were used to visually assess patterns of agreement. The kriging method with the greatest number of highest values of CCC,



kappa, and sensitivity across all variables was determined to be the most optimal kriging method for our data.

#### **4.4 Results**

For both public water supply (Table 4.1) and private well data (Table 4.2), the medians of most parameters, with the exception of manganese, were below the values specified in Saskatchewan's Drinking Water Quality Standards and Objectives. However, the 95<sup>th</sup> percentile value exceeded the standard or objective for several parameters including arsenic, uranium, iron, manganese, sodium, sulphate, alkalinity, hardness and total dissolved solids. When the values for public systems after repeated measures modeling (Table 4.3) were compared to the summary statistics for the raw data (Table 4.1), the median values were similar, but the minimums tended to be larger and the maximums smaller. Although data were available over a 27-year period, the median number of samples by site ranged from 6 to 11 depending on the parameter (Table 4.3).

**Table 4.1** Descriptive statistics for individual samples in public water supplies in Saskatchewan from 1985-2012, along with the Saskatchewan standards and objectives, percent of samples exceeding the respective standard or objective, and percent of samples below detection limit for the water parameters included in the analysis.

	Number of Samples	Mean	Standard deviation	Minimum	25 <sup>th</sup> percentile	Median	75 <sup>th</sup> percentile	95 <sup>th</sup> percentile	Maximum	SK Standard	Exceedances	Below detection limit
	(n)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	% of samples	% of samples
<b>Health standards</b>												
Arsenic	4732	0.0031	0.0067	0.00002	0.00025	0.0009	0.0028	0.014	0.098	0.01	6.9	22.9
Barium	4485	0.047	0.1	0.000025	0.007	0.016	0.046	0.18	2.4	1	0.04	2.8
Boron	4116	0.36	0.5	0.001	0.099	0.24	0.43	1.2	6	5	0.2	5.5
Lead	4569	0.0019	0.0078	0.000005	0.0003	0.0005	0.0025	0.007	0.41	0.01	2.5	67.3
Nitrate	9562	11.57	20.46	0.02	0.5	3	16	42	933	45	4.1	31.4
Selenium	4527	0.0013	0.0056	0.00004	0.0002	0.0005	0.0005	0.004	0.14	0.01	1.9	72.2
Uranium	4617	0.0064	0.011	0.000005	0.0007	0.0028	0.0078	0.023	0.18	0.02	7.3	16.9
<b>Aesthetic objectives</b>										<b>SK Objective</b>		
Alkalinity	5404	408	154	5	308	408	490	674	2451	500	22.0	0
Chloride	5435	48.49	86.91	0.17	9	18	42	233.4	1803	250	4.2	3.8
Copper	4497	0.084	0.29	0.0001	0.006	0.018	0.054	0.31	6.2	1	1.2	6.8
Hardness	4162	536	341	0.5	313	489	740	1107	7800	800	20.7	0.2
Iron	4587	0.30	1.33	0.00025	0.019	0.057	0.21	1.08	46	0.3	18.8	5.9
Magnesium	3120	55.5	37.4	0.5	29.5	49	77	125	449	200	0.002	4.3
Manganese	4614	0.26	1.58	0.00006	0.01	0.065	0.3	0.98	101	0.05	53.5	7.2
Sodium	4353	161.7	189.8	0.5	29	80	221	585	1868	300	18.8	0.1
Sulfate	4284	402.7	366.8	0.1	109.5	326	600	1045	9000	500	32.5	2.5
TDS <sup>1</sup>	4290	1283	661	6.5	761	1199	1725	2453	6687	1500	34.6	0
Zinc	4481	0.028	0.24	0.00013	0.0025	0.007	0.017	0.076	11	5	0.04	25.4

<sup>1</sup>Total Dissolved Solids

**Table 4.2** Descriptive statistics for individual samples in private water supplies in Saskatchewan from 1996-2011, along with the Saskatchewan standards and objectives, percent of samples exceeding the respective standard or objective, and percent of samples below detection limit for the water parameters included in the analysis.

	Number of Samples (n)	Mean (mg/L)	Standard Deviation (mg/L)	Minimum (mg/L)	25 <sup>th</sup> percentile (mg/L)	Median (mg/L)	75 <sup>th</sup> percentile (mg/L)	95 <sup>th</sup> Percentile (mg/L)	Maximum (mg/L)	SK Standard (mg/L)	Exceedances % of samples	Below detection limit % of samples
<b>Health Standards</b>												
Arsenic	4082	0.0050	0.012	0.00005	0.00025	0.0009	0.0043	0.023	0.21	0.01	13.5	21.3
Barium	4082	0.076	0.137	0.00025	0.014	0.032	0.087	0.26	2.19	1	0.4	0.3
Boron	4082	0.33	0.54	0.001	0.068	0.15	0.37	1.4	7.1	5	0.2	1.8
Lead	4082	0.00072	0.0043	0.00005	0.00005	0.0005	0.0005	0.0014	0.21	0.01	0.7	72.9
Nitrate	3996	24.54	73.48	0.02	0.02	1.2	14	126	1300	45	12.2	27.4
Selenium	4076	0.0081	0.036	0.00005	0.0005	0.0005	0.003	0.033	0.84	0.01	11.2	41.1
Uranium	4076	0.012	0.021	0.00005	0.0013	0.0052	0.015	0.044	0.4	0.02	17.8	11.9
<b>Aesthetic Objectives</b>												
Alkalinity	4019	416	148	2	314	399	486	671	1620	500	21.8	n/a <sup>1</sup>
Chloride	4019	69.8	178	0.2	8	21	64	257	4090	250	5.2	1.7
Copper	4080	0.011	0.037	0.0001	0.0005	0.003	0.0094	0.044	1.1	1	0.02	24.1
Hardness	4019	695	569	0.5	335	557	909	1760	6810	800	30.7	0.3
Iron	4091	1.24	2.81	0.00025	0.018	0.12	1.2	6	40	0.3	40.5	1.2
Magnesium	4019	80.96	84.39	0.05	31	60	104	220	1450	200	6.1	0.3
Manganese	4091	0.44	0.68	0.00025	0.023	0.18	0.58	1.7	11	0.05	68.2	2.5
Sodium	4019	181.2	237.3	0.9	25	84	255	653	2710	300	20.8	0
Sulfate	4019	546.3	618.4	0.1	110	354	772	1680	7690	500	39.1	0.15
TDS <sup>2</sup>	4019	1560	1030	6	815	1330	2030	3450	11300	1500	42.7	n/a <sup>1</sup>
Zinc	4081	0.19	1.0	0.00025	0.006	0.018	0.061	0.76	31	5	0.4	15.8

<sup>1</sup>n/a: no detection limit listed. <sup>2</sup>Total Dissolved Solids

**Table 4.3** Descriptive statistics for predicted concentrations after repeated measures modeling for each parameter by site for public water supplies in Saskatchewan, along with Saskatchewan standards and objectives, and the percent of sites with a predicted concentration exceeding the standard or objective. Also shown is a summary of the number of samples per site incorporated in repeated measures models for each parameter.

	Mean predicted concentration											Number of samples per site			
	Number of sites (n)	Mean	SD	Min	Q1	Median	Q3	P95	Max	SK Standard	Exceed	P5	Median	P95	Max
		(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	% of sites	(n)	(n)	(n)
Health Standards															
Arsenic	492	0.0018	0.0029	0.00004	0.0004	0.0008	0.002	0.0074	0.039	0.01	2.0	2	9	18	59
Barium	491	0.037	0.069	0.0002	0.0071	0.015	0.041	0.13	0.74	1	0	2	9	17	28
Boron	477	0.35	0.41	0.0064	0.1	0.23	0.43	1.13	3.03	5	0	1	9	16	26
Lead	491	0.0006	0.00004	0.0005	0.0006	0.0006	0.0007	0.00069	0.00076	0.01	0	2	9	17	28
Nitrate	497	2.67	6.24	0.046	0.62	1.03	2.08	10.42	95.18	45	0.2	2	11	47	366
Selenium	492	0.0006	0.001	0.00008	0.00025	0.0003	0.0005	0.0018	0.012	0.01	0.2	2	9	17	28
Uranium	491	0.0046	0.0063	0.00001	0.00055	0.0026	0.0061	0.016	0.076	0.02	2.6	2	9	18	33
										SK Objective					
Alkalinity	503	400.38	127.3	31.68	316.86	400.10	479.13	612.9	900.4	500	18.1	2	10	20	36
Chloride	499	42.92	64.21	1.88	10.40	19.82	44.08	173.65	489.82	250	2.4	2	10	21	257
Copper	492	0.024	0.025	0.001	0.01	0.015	0.028	0.075	0.17	1	0	1	9	17	28
Hardness	501	491.95	295.6	2.17	298.71	456.52	667.06	1049.18	1482.39	800	15.2	1	8	15	32
Iron	482	0.1	0.12	0.003	0.032	0.06	0.12	0.29	1.18	0.3	4.1	2	9	18	34
Magnesium	483	52.31	34.03	0.58	28.34	46.39	74.16	116.12	191.87	200	0	1	6	14	26
Manganese	483	0.12	0.19	0.00061	0.015	0.059	0.15	0.45	1.84	0.05	52.6	2	9	19	34
Sodium	488	160.3	176.0	2.46	33.11	86.75	232.34	561.17	882.77	300	18.0	1	8	19	34
Sulphate	480	382.8	318.4	0.39	131.59	303.67	553.74	995.08	1930	500	30.0	2	8	37	34
TDS <sup>1</sup>	487	1276.9	590.4	67.75	796.24	1188.64	1691.2	2355.28	3467.45	1500	33.1	1	8	18	34
Zinc	491	0.01	0.0076	0.0019	0.0056	0.0077	0.011	0.022	0.1	5	0	2	9	17	28

SD = standard deviation, Min = minimum, Q1 = 25<sup>th</sup> percentile, Q3 = 75<sup>th</sup> percentile, P5 = 5<sup>th</sup> percentile, P95 = 95<sup>th</sup> percentile, Max = maximum, Exceed = predicted concentration exceeds standard or objective

<sup>1</sup>Total Dissolved Solids

#### 4.4.1 Principal Components Analysis

Bartlett's test of sphericity was satisfied for each dataset indicating that there was sufficient correlation among the variables for PCA to be useful (Table 4.4). The KMO measure of sampling adequacy was above the minimum acceptable value of 0.5 (Kaiser and Rice, 1970) for the data for health standards in private water supplies and aesthetic objectives for both types of supplies (Table 4.4). However, KMO=0.49 for the health standards in the public water supply data. Removing the parameter with the lowest individual measure of sampling adequacy (arsenic) did not substantially improve the KMO. Despite the low KMO, PCA was completed for this dataset to facilitate planned comparisons. However, results of the PCA on health standards in public supplies should be interpreted with caution.

**Table 4.4** Summary of test statistics to assess the adequacy of sampling and correlation among variables in each of the data sets for PCA.

Dataset	Kaiser's measure of sampling adequacy (KMO)	Bartlett's test of sphericity		
		DF	Chi square	P value
<i>Public water supplies</i>				
Health standards	0.4859	21	587.25	<0.0001
Aesthetic objectives	0.6264	55	3705.73	<0.0001
<i>Private wells</i>				
Health standards	0.6203	21	5185.70	<0.0001
Aesthetic objectives	0.6750	55	35445.76	<0.0001

For health standards, analysis of the public supply data yielded three PCs accounting for 63.7 percent of the variance, while the private supply analysis yielded three PCs that accounted for 67.6 percent of the variance (Table 4.5). The first health standards principal component (PC<sub>health</sub>) had strong loadings of nitrate and selenium for both types of supplies. However, the first PC<sub>health</sub> from private supplies had a strong loading of uranium, while in public supplies the loading of uranium was weak and there was also a weak loading of lead. The second PC<sub>health</sub> exhibited opposite pattern loadings, with a strong positive loading of boron and negative loading of barium on PC2<sub>health</sub> in public water supplies, and a strong negative loading of boron and positive loading of barium on PC2<sub>health</sub> in private wells. The third PC<sub>health</sub> had a strong loading of arsenic for the public supplies, but only a weak loading of arsenic in private supplies. For public supplies the

third PC<sub>health</sub> had a moderate loading of uranium, but in the private supplies PC3<sub>health</sub> had a strong loading of lead.

For the aesthetic objectives, PCA on the public supply data yielded four PCs explaining 77.8 percent of the variance, while the private supply data yielded only 3 PCs accounting for 70.3 percent of variance (Table 4.5). For both types of supplies, the first aesthetic objectives PC (PC<sub>aesthetic</sub>) had moderate to strong loadings of alkalinity, chloride, sodium, sulphate, and total dissolved solids. Also, for both types of supplies, the second PC<sub>aesthetic</sub> was characterized by strong contributions from hardness and magnesium with weak contributions by sulphate and manganese. In addition, private supplies had weak loadings of zinc on PC2<sub>aesthetic</sub>. The third PC<sub>aesthetic</sub> for public and private supplies were both characterized by contributions from iron and manganese, but in private systems PC3<sub>aesthetic</sub> also exhibited a negative loading of copper. Conversely, the PC4<sub>aesthetic</sub> retained only from the public supply data had moderately strong positive loadings of copper and zinc.

**Table 4.5** Principal components analysis on public water supplies and private well water supplies: varimax rotated principal components patterns, and eigenvalues and percent of variance explained for each retained component. The loadings in bold font are the maximum loading for each variable.

	<b>Public Water Supplies</b>				<b>Private Wells</b>		
<i>Health Standards</i>	PC1	PC2	PC3		PC1	PC2	PC3
Arsenic	-0.121	0.142	<b>0.808</b>		-0.341	-0.091	<b>0.474</b>
Barium	0.047	<b>-0.818</b>	-0.141		-0.041	<b>0.893</b>	0.100
Boron	-0.062	<b>0.903</b>	-0.123		-0.195	<b>-0.818</b>	0.168
Lead	<b>0.472</b>	0.092	0.171		0.156	0.026	<b>0.893</b>
Nitrate	<b>0.768</b>	-0.071	-0.164		<b>0.770</b>	0.275	-0.110
Selenium	<b>0.867</b>	-0.220	0.019		<b>0.853</b>	-0.007	0.074
Uranium	0.387	-0.290	<b>0.576</b>		<b>0.772</b>	-0.013	-0.049
Eigenvalue	2.127	1.275	1.059		2.290	1.381	1.057
Cumulative variance (%)	30.4	48.6	63.7		32.7	52.5	67.6
<i>Aesthetic Objectives</i>	PC1	PC2	PC3	PC4	PC1	PC2	PC3
Alkalinity	<b>0.755</b>	0.111	0.164	-0.198	<b>0.687</b>	-0.023	0.217
Chloride	<b>0.753</b>	-0.193	0.002	0.226	<b>0.779</b>	0.043	-0.195
Copper	0.127	0.012	-0.200	<b>0.714</b>	0.030	0.223	<b>-0.757</b>
Hardness	0.009	<b>0.973</b>	0.066	0.042	0.067	<b>0.960</b>	0.038
Iron	0.138	-0.089	<b>0.901</b>	0.053	0.117	0.121	<b>0.784</b>
Magnesium	-0.014	<b>0.961</b>	0.055	0.038	0.103	<b>0.951</b>	0.020
Manganese	0.188	0.452	<b>0.711</b>	-0.065	0.062	0.468	<b>0.663</b>
Sodium	<b>0.914</b>	-0.199	0.136	0.026	<b>0.922</b>	-0.116	0.118
Sulphate	<b>0.663</b>	0.517	0.018	0.116	<b>0.609</b>	0.555	0.076
Total Dissolved Solids	<b>0.920</b>	0.288	0.121	-0.016	<b>0.907</b>	0.325	0.082
Zinc	-0.089	0.078	0.237	<b>0.763</b>	-0.091	<b>0.396</b>	-0.375
Eigenvalue	3.746	2.362	1.264	1.181	3.775	2.184	1.779
Cumulative variance (%)	34.1	55.5	67.0	77.8	34.3	54.2	70.4

PC = principal component

#### *4.4.2 Geostatistical Analysis*

Across Saskatchewan, there were data available for 492 groundwater-sourced public water supplies. Of these, 480 fell within the study area outlined in Figure 4.1 and were used in the geostatistical analysis. Arsenic concentrations were available from all 480 public water supply locations within the study area. Health standards PC scores were available for 459 locations and aesthetic objectives for 435 locations. Data was available for 4093 private wells in total; 4084 of the private well locations fell within the study area and were used in the geostatistical analysis. Arsenic concentrations were available for 4073 private wells within the study area, health standards PC scores were available for 3970 private wells, and aesthetic objectives for 3999 private wells in the study area.

The spherical semivariogram models had the lowest AIC for most variables; for the variables where the spherical semivariogram did not have the lowest AIC, there was little difference in AICs among all models (Table 4.6). To maintain consistency, the spherical model was used for all variables for ordinary and universal kriging. Because the spherical variogram model is not an option for empirical Bayesian kriging, the default setting for the power variogram was used. Visual inspection for anisotropy suggested it was mild, if present; therefore isotropic models were assumed for all variables. Visual assessment of Q-Q plots suggested that deviations from normality were mild. Mild to moderate non-stationarity was evident on examination of Voronoi maps of local means; detrending was applied to the data prior to kriging.



**Table 4.6** Summary of variogram models for each variable in public and private water supplies. Inputs into models were lag distance (Average nearest neighbor distance was used) and number of lags. Large scale trends identified in the data are reported, along with results of variogram modeling (model form, Moran's I, Geary's C, nugget, and range).

Variable	N <sup>1</sup>	Lag distance (km)	Number of lags	Large scale trend	Model	Moran's I (p value)	Geary's C (p value)	Nugget	Range (km)
<i>Public supplies</i>									
Arsenic	336	15.28	25	2 <sup>nd</sup> order	Spherical	0.34 (<0.001)	0.76 (0.02)	1.01	190.9
PC1 <sub>health</sub>	321	16.40	24	1 <sup>st</sup> order	Spherical	0.11 (0.15)	1.09 (0.47)	0.92	196.7
PC2 <sub>health</sub>	321	16.40	24	1 <sup>st</sup> order	Spherical	0.15 (0.05)	0.79 (0.03)	0.86	196.7
PC3 <sub>health</sub>	321	16.40	24	1 <sup>st</sup> order	Spherical	0.34 (<0.001)	0.63 (<0.001)	0.62	196.7
PC1 <sub>aesthetic</sub>	305	16.31	24	1 <sup>st</sup> order	Spherical	0.16 (0.05)	0.75 (0.01)	0.8	195.7
PC2 <sub>aesthetic</sub>	305	16.31	24	2 <sup>nd</sup> order	Spherical	0.27 (0.001)	0.67 (0.004)	0.56	195.7
PC3 <sub>aesthetic</sub>	305	16.31	24	2 <sup>nd</sup> order	Spherical	0.11 (0.19)	1.018 (0.85)	0.92	195.7
PC4 <sub>aesthetic</sub>	305	16.31	24	none	Gaussian	-0.02 (0.89)	0.91 (0.36)	0.96	195.7
<i>Private supplies</i>									
Arsenic	2851	3.31	91	1 <sup>st</sup> order	Spherical	0.14 (<0.001)	0.90 (0.06)	2.39	150.9
PC1 <sub>health</sub>	2779	3.30	91	2 <sup>nd</sup> order	Exponential	0.23 (<0.001)	0.69 (<0.001)	0.99	151.1
PC2 <sub>health</sub>	2779	3.30	91	2 <sup>nd</sup> order	Spherical	0.19 (<0.001)	0.45 (<0.001)	0.72	151.1
PC3 <sub>health</sub>	2779	3.30	91	2 <sup>nd</sup> order	Gaussian	0.12 (<0.001)	1.12 (0.12)	0.94	151.1
PC1 <sub>aesthetic</sub>	2799	3.23	95	2 <sup>nd</sup> order	Spherical	0.26 (<0.001)	0.43 (<0.001)	0.84	153.4
PC2 <sub>aesthetic</sub>	2799	3.23	95	2 <sup>nd</sup> order	Spherical	0.21 (<0.001)	0.41 (<0.001)	0.78	153.4
PC3 <sub>aesthetic</sub>	2799	3.23	95	1 <sup>st</sup> order	Spherical	0.10 (<0.001)	0.84 (0.01)	0.96	153.4

N<sup>1</sup> = number of sites used in training dataset (70% of sites within study area)

PC = principal component

The concordance between the observed values in the validation data and the values predicted from kriging (Table 4.7) for the public water supply data ranged from 0.01 for universal kriging of arsenic concentrations to 0.35 for universal kriging of the second PC2<sub>aesthetic</sub>. For the private well data, Lin's CCC ranged from -0.001 for universal kriging of PC1<sub>health</sub> to 0.52 for ordinary kriging of PC1<sub>aesthetic</sub> and empirical Bayesian kriging of PC2<sub>aesthetic</sub> and PC2<sub>health</sub>.

**Table 4.7** Values of Lin's concordance correlation coefficient comparing values in the validation datasets and values predicted by ordinary kriging, universal kriging, and empirical Bayesian kriging for log arsenic concentration and component scores for public and private water supplies.

	Ordinary		Universal		Empirical Bayesian	
	rho	p value	rho	p value	rho	p value
<b><i>Public supplies</i></b>						
Arsenic	0.17	0.01	0.01	0.87	0.17	0.008
PC1 <sub>health</sub>	0.08	0.008	0.12	<0.001	0.12	0.007
PC2 <sub>health</sub>	0.17	<0.001	0.19	<0.001	0.20	0.001
PC3 <sub>health</sub>	0.15	0.007	0.15	0.015	0.17	0.009
PC1 <sub>aesthetic</sub>	0.18	<0.001	0.20	<0.001	0.23	<0.001
PC2 <sub>aesthetic</sub>	0.28	<0.001	0.35	<0.001	0.24	<0.001
PC3 <sub>aesthetic</sub>	0.10	0.02	0.04	0.62	0.18	<0.001
PC4 <sub>aesthetic</sub>	0.03	0.14	0.03	0.14	0.08	0.06
<b><i>Private wells</i></b>						
Arsenic	0.16	<0.001	0.16	<0.001	0.23	<0.001
PC1 <sub>health</sub>	0.32	<0.001	-0.001	0.38	0.31	<0.001
PC2 <sub>health</sub>	0.45	<0.001	0.48	<0.001	0.52	<0.001
PC3 <sub>health</sub>	0.20	<0.001	0.18	<0.001	0.22	<0.001
PC1 <sub>aesthetic</sub>	0.52	<0.001	0.00	0.55	0.49	<0.001
PC2 <sub>aesthetic</sub>	0.40	<0.001	0.43	<0.001	0.52	<0.001
PC3 <sub>aesthetic</sub>	0.19	<0.001	0.18	<0.001	0.19	<0.001

PC = principal component

Spearman's rank correlation coefficients (Table 4.8) summarizing the correlation between observed values in the validation data and the values predicted from kriging ranged from 0.006 to 0.38 for the public water supply data and from 0.24 to 0.53 for the private supply data (Table 4.8).

**Table 4.8** Values of Spearman's rank coefficient for correlation between actual values and values predicted by ordinary kriging, universal kriging, and empirical Bayesian kriging for log arsenic concentration and component scores for public and private water supplies.

	Ordinary		Universal		Empirical Bayesian	
	rho	p value	rho	p value	rho	p value
<b><i>Public supplies</i></b>						
Arsenic	0.19	0.02	0.006	0.94	0.19	0.02
PC1 <sub>health</sub>	0.20	0.02	0.25	0.003	0.19	0.03
PC2 <sub>health</sub>	0.29	<0.001	0.28	0.001	0.25	0.003
PC3 <sub>health</sub>	0.21	0.01	0.23	0.007	0.21	0.01
PC1 <sub>aesthetic</sub>	0.28	0.001	0.28	0.001	0.31	<0.001
PC2 <sub>aesthetic</sub>	0.38	<0.001	0.40	<0.001	0.33	<0.001
PC3 <sub>aesthetic</sub>	0.22	0.01	0.31	<0.001	0.30	0.001
PC4 <sub>aesthetic</sub>	0.09	0.33	0.09	0.33	0.15	0.1
<b><i>Private wells</i></b>						
Arsenic	0.24	<0.001	0.24	<0.001	0.28	<0.001
PC1 <sub>health</sub>	0.37	<0.001	0.24	<0.001	0.36	<0.001
PC2 <sub>health</sub>	0.49	<0.001	0.49	<0.001	0.52	<0.001
PC3 <sub>health</sub>	0.28	<0.001	0.26	<0.001	0.29	<0.001
PC1 <sub>aesthetic</sub>	0.53	<0.001	0.45	<0.001	0.51	<0.001
PC2 <sub>aesthetic</sub>	0.40	<0.001	0.41	<0.001	0.50	<0.001
PC3 <sub>aesthetic</sub>	0.27	<0.001	0.25	<0.001	0.24	<0.001

The value of the kappa statistic summarizing agreement between observed values in the validation data and the values predicted from kriging classified above the 75<sup>th</sup> percentiles for the observed data ranged from -0.08 to 0.19 for the public water supply data and 0.0 to 0.39 for the private supply data (Table 4.9). In several cases, kappa=0 and p values were not produced because no predicted values were greater than the observed data's 75<sup>th</sup> percentile (Table 4.9).

Kappa was calculated for whether the predicted concentration of arsenic was greater than the drinking water standard of 0.01 mg/L, but predictions above the drinking water standard were only obtained for universal kriging for the public supply data (kappa = 0.16, p=0.02) and empirical Bayesian kriging for the private well data (kappa=0.03, p=0.006).

The sensitivity for the different kriging methods at predicting values above the 75<sup>th</sup> percentile ranged from 0.0 to 0.33 for the public water supply data, and 0.0 to 0.38 for the private supply

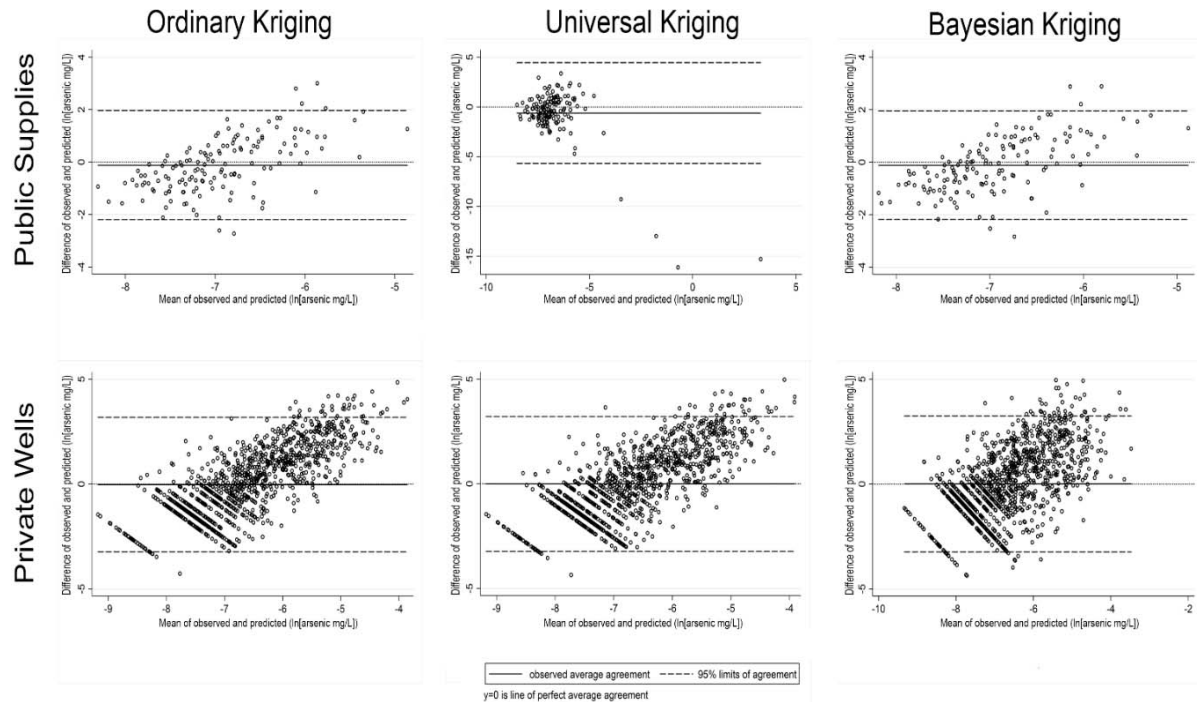
data (Table 4.9). Specificity was also calculated and ranged from 0.73 to 1.0 for the public supply data and 0.88 to 1.0 for the private well data. Overall, empirical Bayesian kriging was most likely to rank above the other methods when comparing CCC, rho, Kappa and sensitivity across all variables considering both public and private supplies, and was selected as the best of the methods evaluated for these data.

**Table 4.9** Values of the kappa statistic for agreement between values predicted by ordinary, universal and empirical Bayesian kriging and measured values being over the 75<sup>th</sup> percentile for observed arsenic concentrations and component scores for public and private water supplies. Also included are kappa statistics for being over the drinking water standard (0.01 mg/L) for arsenic, and sensitivity of the various kriging techniques for identifying predicted values over the 75<sup>th</sup> percentile of the observed data.

	Ordinary Kriging				Universal Kriging				Empirical Bayesian Kriging			
	Kappa	p	Se	Sp	Kappa	p	Se	Sp	Kappa	p	Se	Sp
<i>Public supplies</i>												
As > 10 µg/L	0	n/a	0.00	1.00	0.16	0.02	0.33	0.95	0	n/a	0	1.00
As > 75th percentile	0.09	0.13	0.19	0.88	0.06	0.23	0.33	0.73	0.19	0.008	0.25	0.91
PC1 <sub>health</sub>	0	n/a	0	1.00	0	n/a	0	1.00	0.13	0.009	0.11	0.98
PC2 <sub>health</sub>	0	n/a	0	1.00	-0.01	0.72	0	0.99	0.07	0.13	0.09	0.96
PC3 <sub>health</sub>	0.09	0.03	0.09	0.98	0.09	0.03	0.09	0.98	0.06	0.19	0.11	0.93
PC1 <sub>aesthetic</sub>	0.03	0.21	0.03	0.99	0.04	0.22	0.06	0.97	0.16	0.02	0.21	0.92
PC2 <sub>aesthetic</sub>	0.15	0.01	0.16	0.96	0.18	0.008	0.19	0.95	0.16	0.02	0.19	0.94
PC3 <sub>aesthetic</sub>	-0.08	0.87	0.03	0.91	-0.03	0.65	0.13	0.85	-0.009	0.54	0.13	0.87
PC4 <sub>aesthetic</sub>	0	n/a	0	1.00	0	n/a	0	1.00	-0.02	0.72	0	0.99
<i>Private wells</i>												
As > 10 µg/L	0	n/a	0	1.00	0	n/a	0	1.00	0.03	0.006	0.03	0.99
As > 75th percentile	0.03	.04	0.04	0.98	0.01	0.27	0.04	0.97	0.12	<0.001	0.16	0.93
PC1 <sub>health</sub>	0.18	<0.001	0.26	0.95	0.2	<0.001	0.30	0.88	0.13	<0.001	0.13	0.97
PC2 <sub>health</sub>	0.29	<0.001	0.26	0.97	0.35	<0.001	0.35	0.94	0.39	<0.001	0.36	0.96
PC3 <sub>health</sub>	0.04	0.02	0.06	0.97	0.1	<0.001	0.18	0.90	0.07	0.002	0.10	0.95
PC1 <sub>aesthetic</sub>	0.27	<0.001	0.26	0.95	0.3	<0.001	0.38	0.89	0.24	<0.001	0.24	0.95
PC2 <sub>aesthetic</sub>	0.14	<0.001	0.15	0.96	0.24	<0.001	0.32	0.89	0.24	<0.001	0.29	0.91
PC3 <sub>aesthetic</sub>	0.01	0.12	0.02	0.99	0.01	0.16	0.03	0.98	0.02	0.04	0.02	0.99

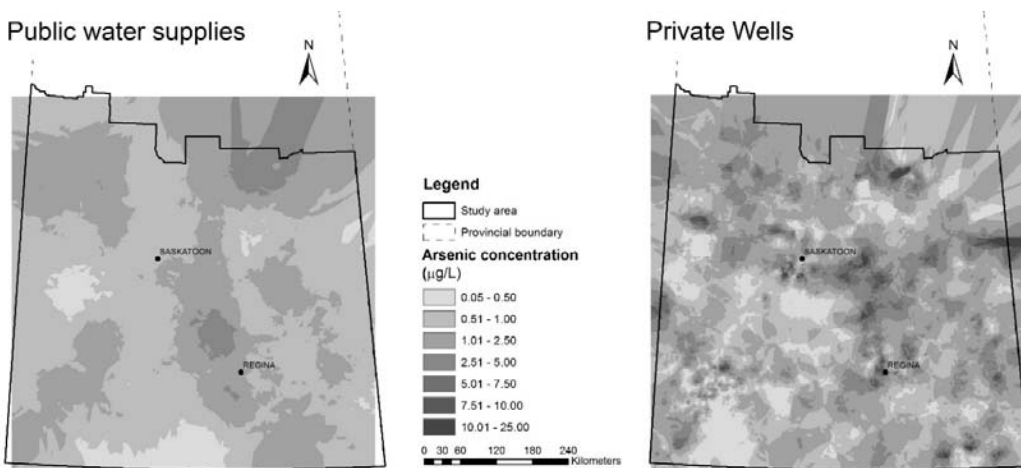
Se = sensitivity, SP = specificity, As = arsenic, n/a = p value not obtained due to occurrence of zero cells in contingency table (no values predicted above cut off value).

Limits of agreement plots for arsenic for public supply and private well data for the various kriging methods (Figure 4.2) illustrate that at higher concentrations kriging tended to underestimate predicted values, and where concentrations were very low, kriging tended to overestimate the predicted values. However, this pattern was reversed for universal kriging for the public supply data. Universal kriging resulted in some extremely high predicted arsenic values from the public data, suggesting this method was unsuitable for this data.



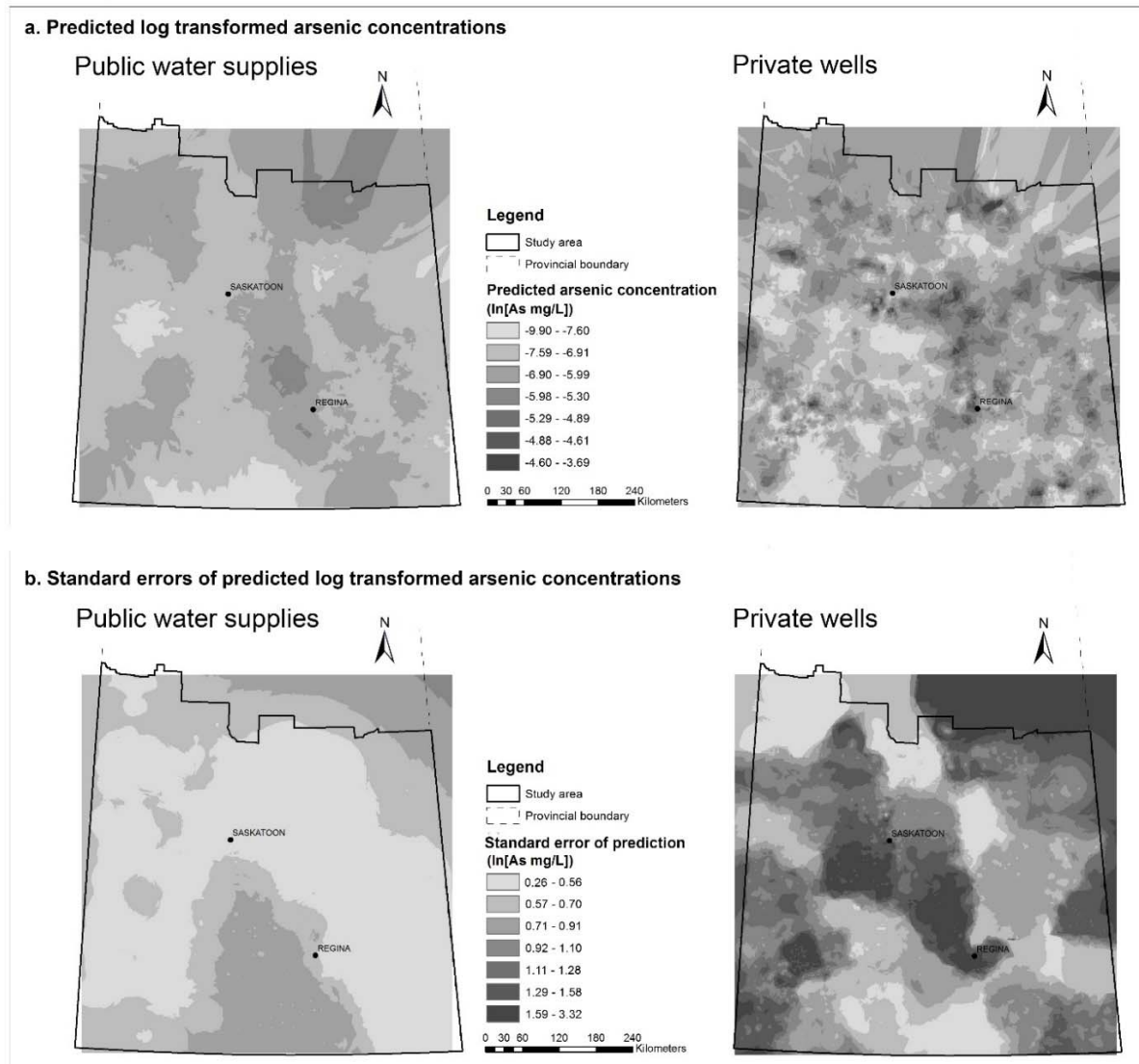
**Figure 4.2** Limits of agreement plots for between observed and predicted  $\ln(\text{arsenic concentrations in mg/L})$  for the different types of kriging for public supplies (top row) and private supplies (bottom row).

Results from empirical Bayesian kriging of log transformed arsenic concentrations from the public and private water supply data within the study area were back transformed to obtain surfaces of predicted arsenic concentrations (Figure 4.3).



**Figure 4.3** Arsenic concentrations ( $\mu\text{g/L}$ ) predicted by empirical Bayesian kriging for public water supply data (left) and private well data (right) in study area in southern Saskatchewan.

Prediction surfaces based on the natural logarithm transformed arsenic concentrations are shown in Figure 4.4a, along with their corresponding prediction standard error surfaces which illustrate the amount of uncertainty associated with the predicted values (Figure 4.4b).



**Figure 4.4** Empirical Bayesian kriging prediction maps for natural logarithm transformed arsenic concentrations (a) along with the prediction standard error maps (b) for public supplies (on left) and private supplies (on right) in study area in southern Saskatchewan.

## 4.5 Discussion

In Saskatchewan, groundwater sources are primarily utilized by residents of rural and remote areas (Environment and Climate Change Canada, 2007). There were a considerable number of water samples that exceeded drinking water standards and objectives in the surveillance data, highlighting the need to promote adequate testing of drinking water in rural areas. Generally, it appears that contaminants listed as aesthetic objectives exceed guideline values at a higher frequency than the health-related standards. As expected, the raw groundwater sampled from private wells exceeded standards and objectives more frequently than the water from the regulated, treated public water supplies. However, there were still a considerable number of samples from public supplies that exceeded guidelines, especially for aesthetic objectives.

In a previous study, a sample of 283 wells in Saskatchewan (Thompson, 2003) found that approximately 45 percent of the wells exceeded the Saskatchewan drinking water objective for sulphate, 47 percent exceeded the objective for iron, 61 percent exceeded the objective for hardness, and 79 percent exceeded the objective for manganese. Our study reflected a similar pattern, although the rates of exceedances were slightly lower; approximately 39 percent for sulphate, 40 percent for iron, 31 percent for hardness, and 68 percent for manganese in the private wells, and 32 percent for sulphate, 19 percent for iron, 21 percent for hardness and 53 percent for manganese in public supplies. A previous study from Saskatchewan reported that the aesthetic quality of tap water is linked to safety concerns; the perception that tap water was unsafe was associated by survey participants with their assessment of its aesthetic qualities (McLeod et al., 2015). The aesthetic quality of tap water could act as a determinant of health by increasing consumption of water alternatives which may include sugar-sweetened beverages (Onufrak et al., 2014).

Previous studies have also investigated concentrations of arsenic (Thompson et al., 1999) and nitrate (Thompson, 2001) in Saskatchewan wells. Thompson et al. (1999) sampled 61 wells (private wells and wells maintained by rural municipalities) for arsenic, and found that 23 percent exceeded the current Saskatchewan drinking water standard applied to regulated public water supplies. In our study, just over 13 percent of private wells exceeded the standard, while approximately 7% of public supply samples exceeded the standard. Thompson (2001) found that 14 percent of wells tested exceeded the standard for nitrate for WSA regulated waterworks,



while 12 percent of the private wells included in our study exceeded the standard. However, only 4 percent of public supplies exceeded the nitrate standard.

#### *4.5.1 Principal Components Analysis*

PCA has been used in previous studies to examine and interpret patterns of groundwater quality parameters (Helena et al., 2000; Sánchez-Martos et al., 2001; Liu et al., 2003; Chapagain et al., 2010; Singh et al., 2010; Belkhiri et al., 2011). These types of studies typically identify common factor patterns and interpret them with respect to presumed natural and anthropogenic processes that impact groundwater quality, and are often focused on major ions (e.g., sodium, chloride, magnesium, sulphate) that would fall under aesthetic objectives in the Saskatchewan Drinking Water Quality Standards and Objectives. PCA analysis of groundwater has often included nitrate, which falls under Saskatchewan health standards, as a marker for anthropogenic influences on groundwater (e.g., Helena et al., 2000; Sánchez-Martos et al., 2001; Liu et al., 2003; Chapagain et al., 2010). However, the full range of parameters included in, and among such studies has not been consistent particularly with respect to the inclusion of trace metals, making it somewhat difficult to compare results. Comparison to our study was further hampered because we analyzed health standards and aesthetic objectives separately to align our analysis with Saskatchewan Drinking Water Quality Standards and Objectives.

We limited our analyses to include parameters that were routinely sampled from both the public and private supply data to facilitate comparison between the differing supplies. We expected the results to differ between the types of systems because the public supply data represents treated water supplies and the private water supply data represents raw well water samples. While there were some differences in the principal components extracted from the public and private data, there were some striking similarities. This was especially true in the results for the aesthetic objectives, even though a four  $PC_{\text{aesthetic}}$  were retained for the public supply data and three for the private well data. The first  $PC_{\text{aesthetic}}$  was associated with the same group of variables in both data sets: sodium, chloride, sulphate, alkalinity and total dissolved solids. Additionally, hardness and magnesium were strongly associated with the second PC and iron and manganese with the third PC in both public and private water supplies. The consistent loading of these parameters between the data sets suggest relatively strong associations between these parameters in Saskatchewan groundwater.

The PCA for health standards also exhibited some consistencies: nitrate and selenium were strongly associated with the first PC<sub>1health</sub> for both public and private supplies. Arsenic was associated with the third PC<sub>health</sub> in both data sets, but strongly associated with that PC in the public supply data. In contrast, Uranium was associated with PC<sub>3health</sub> in public supplies and with PC<sub>1health</sub> in private wells. In addition, lead was associated with PC<sub>1health</sub> in public supplies, and with PC<sub>3health</sub> in private wells. Because lead contamination of water can be associated with leaching from distribution systems, differences in the covariance of lead with other parameters between public supplies and private wells is not unexpected. However, caution is warranted in the interpretation of the PCA for health standards from the public supply data considering the low Kaiser's measure of sampling adequacy for these data.

The retained principal components for the health-related parameters explained less variability than those for the aesthetic objectives for both the public supply and private well data, suggesting that these methods might be expected to work better for parameters associated with the aesthetic qualities of water delineated in the Saskatchewan Drinking Water Quality Standards and Objectives. It is possible that the parameters associated with aesthetic qualities are more closely correlated than parameters associated with health risks.

The parameters associated with aesthetic objectives are also found in relatively high concentrations in Saskatchewan groundwater, whereas the contaminants included in the health standards were often found at low concentrations and often below detection limits. Although we excluded some parameters (e.g., cadmium, chromium) due to very high proportions of non-detects, some of the other variables retained in the analysis had levels of non-detects that might impair their performance in PCA. Farnham et al. (2002) suggest that when non-detects comprise greater than 25% of samples, a variable's performance in PCA deteriorates. The variables for which results were consistent between datasets typically had a low proportion of non-detects in the data.

#### *4.5.2 Geostatistical Analysis*

Kriging has previously been validated as a method to summarize arsenic concentrations in groundwater quality and in one study was found to be superior to using an area average or nearest well as a proxy to predict well concentrations (Meliker et al., 2008). While some studies have investigated the use of indicator kriging to model the probability of higher arsenic

concentrations using geological and hydrological covariates (Goovaerts et al., 2005; Yang et al., 2009), some recent studies have compared various kriging methods that are accessible in GIS software to investigate prediction of arsenic concentrations in groundwater (Gong et al., 2014; James et al., 2014). James et al. (2014) evaluated the performance of various kriging methods (ordinary, universal, simple kriging with varying means, kriging with external drift, cokriging with ordinary kriging and cokriging with universal kriging) over a relatively small area in Colorado and found that ordinary kriging performed best. Gong et al. (2014) compared inverse distance weighted interpolation with kriging using Gaussian and spherical models as well as cokriging in predicting arsenic concentrations over various regions in Texas, and found regional differences in the performance of kriging, and concluded that kriging over smaller areas was more accurate than over large geographic regions.

Studies done in different areas over different scales using different methodologies and different covariates make comparisons difficult, but it is apparent that the performance of kriging in our study generally predicted arsenic concentrations poorly compared to other studies. Although we found that empirical Bayesian Kriging performed the best overall in our study, agreement between our predicted values and the validation datasets was weak. However, methods for assessing agreement vary among studies. Rather than relying only on Pearson's or Spearman's rank correlations to assess agreement, we chose Lin's CCC which considers the scale of variables along with their relationship so is arguably a better measure of agreement than correlation. For arsenic, our values of CCC indicate weak agreement across methods and data sets; however, this measure hasn't been used in previous studies. A study based in Colorado reported a kappa of 0.58 for the prediction of arsenic concentration above the drinking water standard of 0.01 mg/L (James et al., 2014) whereas in our study kappa was just 0.02. Each of our methods also had a very low sensitivity at detecting higher levels of arsenic in groundwater. In contrast, James et al. (2014) reported a sensitivity of 100% at estimating whether a well has a concentration of arsenic greater than 0.01 mg/L while Gong et al. (2014) reported a sensitivity of 89 percent in a subset of samples, compared to our sensitivity of 2 percent (empirical Bayesian kriging, private wells). However, using indicator kriging to predict the presence of arsenic in wells at concentrations greater than 0.5 µg/L, (Ayotte et al., 2006) reported a sensitivity of 37 percent.

From the limits of agreement plots for arsenic, it is apparent that empirical Bayesian kriging tended to underestimate high values of arsenic and overestimate very low values, which explains the low sensitivity at predicting elevated concentrations of arsenic. However, it does predict some areas with relatively high arsenic concentrations compared to the rest of Saskatchewan. Based on a cluster of wells with arsenic concentrations above the drinking water standard, a previous study identified a putative arsenic hotspot (Thompson et al., 1999) that falls within a relatively large area of higher predicted arsenic concentrations in the northeast quadrant of our study area. While our values for Lin's CCC were generally quite low, the associated p values do indicate a significant relationship, particularly for empirical Bayesian kriging. The Spearman's rank correlation coefficients tended to be larger than Lin's CCC, which suggested that although the predicted concentrations did not agree in magnitude with the observed concentrations, they did reasonably reflect where relatively higher and lower values might be expected.

We elected not to use covariate information such as well depths or geological data in our models due to difficulty in obtaining accurate covariate data over our study area. While depth might be expected to improve modeling of arsenic concentration, conflicting results from other studies suggest that the contribution of depth may be dependent on the study area. For example, a negative correlation between increasing well depth and arsenic concentrations has been reported for wells in Bangladesh (Hassan and Atkins, 2011; Yu et al., 2003), while a positive association between well depth and arsenic was reported in North Carolina (Kim et al., 2011). Yang et al. (2009) did not detect any association between arsenic concentration and well depth in Maine. In one study, including well depth in cokriging models did not improve the ability of kriging to predict arsenic levels (James et al., 2014). Gong et al. (2014) found that incorporating well depth in cokriging did not necessarily improve the correlation between predicted and actual values, but did improve the performance of regression models used to predict arsenic levels. Furthermore, Yu et al. (2003) investigated factors affecting arsenic at different geographic scales and concluded that much of the variability in arsenic concentrations at a scale of less than 3 km could be explained by well depth, while geology was the most important factor at scales of greater than 10 km. This suggests that given the large scale of our study area relative to other reported studies, it is unlikely that adding well depth as a covariate would have improved our models.

While incorporation of geological data might have improved our predictions, this information was not available for the large study area.

Others have reported a tremendous amount of heterogeneity in groundwater concentrations of arsenic over small scales that is poorly understood (Yu et al., 2003; Yang et al., 2009). In Bangladesh, wells within a radius of less than 1 km were found to vary by up to 1000 µg/L (Yu et al., 2003). In another study of a relatively small region of Bangladesh, wells in close proximity exhibited extremely variable arsenic concentrations, especially wells less than 30 m in depth (Van Geen et al., 2003). This issue was also highlighted in the geostatistical analysis of arsenic in wells in Michigan; residuals for predicted arsenic values were mapped and no spatial pattern in the residuals was detected (Meliker et al., 2008). The close proximity of wells with negative and positive residuals of greater than 10 µg/L reflected high variability in arsenic concentrations over short distances (Meliker et al., 2008). Additionally, a study in Texas compared geostatistical methods among regions, and found the performance of the different methods varied less within a given area than across the different regions (Gong et al., 2014). This suggests that variability in the distribution of groundwater arsenic across regions is a limiting factor in identifying a single method that would perform uniformly well in different geographic areas. Given the apparent differences in processes influencing spatial variability of arsenic at different scales, it is possible that developing kriging models over smaller targeted areas with a high density of samples could have improved the performance of our predictions for some local regions.

Interpretation of mapped results of PCA scores is less straightforward because the values are a representation of a combination of parameters that contribute to the PCA components. For example, areas with high values for the objectives principal component one represent higher predicted concentrations of one or more of the contributors to this component, including sodium, chloride, sulphate, alkalinity and total dissolved solids. Nevertheless, this method is useful for examining patterns in common grouping of parameters and allowed extraction of factor scores to summarize mixtures of variables over geographic regions for use in other analyses.

Previous studies have used geostatistical methods to map the scores resulting from PCA or FA and used the resultant maps to predict the factors that may be impacting groundwater quality, such as pollution or salt water intrusion (Sánchez-Martos et al., 2001; Satyaji Rao et al., 2009; Shyu et al., 2011; Nazzal et al., 2015). It does not appear that the use of kriging with PCA or

factor analysis has been well-validated for prediction of groundwater quality. We are not aware of other studies that have assessed the ability of kriging to accurately predict PCA scores at unmeasured locations so we cannot compare our results to others. For our data, agreement between predicted principal component scores and actual scores in the validation datasets was generally quite low.

Given that we know that the performance of kriging at predicting arsenic concentrations over our study area was not strong, we could reasonably expect to see a further loss of predictive ability by combining a variable reduction method such as PCA with kriging. PCA reduces the dimensionality of a dataset while capturing as much of the information in the original variables as possible. In our data the percentage of variance retained by the PC ranged from 63.7 to 77.8%. However, the predictive performance of factor scores, as determined by our measures of agreement, generally did not appear to be substantially different than the prediction of arsenic concentrations. In some cases we achieved better agreement between predicted and calculated PC scores than with arsenic concentrations.

While spatial patterns of arsenic have been studied extensively, spatial patterns of the other variables and especially mixtures of variables have not. Therefore, it is possible that the PCs we extracted are subject to variability at scales not captured by our analysis. The use of PCA combined with kriging of factor scores shouldn't be discounted as a means of summarizing water quality but should be interpreted with caution given the weak agreement between predicted and actual PC scores in our study, and a lack of other studies validating geostatistical analysis of PC scores.

#### *4.5.3 Limitations*

It is estimated that there are over 66,000 wells in Saskatchewan (Thompson et al., 1999) and our sample of 4093 private wells is a non-random sample of less than 10 percent of privately owned wells in Saskatchewan. Because the database consists of samples taken through participation in a voluntary water quality program, it could disproportionately represent residents with concerns about their well water quality.

Although the public supply data represents data from all available water supplies across Saskatchewan, there were relatively few locations represented in the public supply data relative

to the size of the study area, resulting in a low sampling density that may have particularly impacted the ability of kriging to capture the variability of arsenic at small spatial scales.

The variability in the public well data was also attenuated by the use of repeated measures modeling to estimate a single predicted value for each parameter at each site. While spatio-temporal kriging methods are available in some software packages (e.g. Gräler et al., 2016), we elected to not incorporate temporal analysis. The inclusion of time had a minimal effect on the predicted concentrations in all the repeated measures models (results not reported), and data were available at very irregular intervals and for a small proportion of years for the majority of sampling sites. Calculating PCA at different time points for spatio-temporal analysis would have been extremely complicated. Furthermore, the ultimate goal of this analysis was to estimate an average exposure over multiple years for an epidemiological analysis of associations between water quality and chronic diseases for which the relevant induction period is uncertain (Chapters 5 and 6). Therefore, it was decided the estimation of a mean value for each of the public water supplies represented the most parsimonious approach.

Our PCA may have been hampered by not being able to make use of a full suite of parameters especially with respect to the health-related standards. We also made the decision to separately analyze aesthetic and health parameters because they are segregated in drinking water standards and objectives. It is possible that considering all available parameters together could have improved the performance of the PCA, although it seems likely that the high number of samples below detection limits would continue to limit the usefulness of some of the variables measured as health standards.

Kriging methods rely on an estimation of the spatial structure of data. While semivariogram models provide a means of investigating spatial relationships, kriging typically requires the assumption that the chosen semivariogram model represents the true spatial structure. This assumption is relaxed with the empirical Bayesian kriging methods which allows for uncertainty in the semivariogram parameters which likely contributes to the superior predictive performance of this method in our study. However, other researchers have investigated Bayesian statistical methods to predict arsenic groundwater concentrations which incorporate spatial relationships using alternatives to semivariograms (Kim et al., 2011; Sanders et al., 2012). Use of methods such as these could potentially be used to improve prediction of arsenic concentrations and

overcome some of the limitations of kriging especially when spatial variability arises from different processes at different scales limiting the effectiveness of variogram modeling, even after allowing for uncertainty in the semivariogram.

#### **4.6 Conclusion**

In this study we investigated the use of kriging to predict groundwater concentrations of arsenic across southern Saskatchewan. We also investigated the use of PCA to summarize health standards and aesthetic objectives, as defined in the Saskatchewan Drinking Water Quality Standards and Objectives, and used kriging to summarize the results of the PCA across the same region of Saskatchewan. We compared ordinary, universal and Bayesian kriging for predicting log arsenic concentrations and PC scores across the study area for public and private water supplies. Across all the variables investigated, Bayesian kriging resulted in the best agreement between predicted and actual values in a validation dataset. However, there was only weak to moderate agreement between predicted and actual values, limiting the effectiveness of kriging to estimate values for arsenic concentrations or PC scores across our large study area. The methods examined were not sensitive for identifying arsenic concentrations above the drinking water standard, nor predicting values in the 75<sup>th</sup> percentile of the validation dataset. Therefore, Bayesian kriging across large areas of Saskatchewan cannot be considered an optimal method for predicting actual arsenic concentrations or PC scores representing combinations of other parameters making up the Saskatchewan Drinking Water Quality Standards and Objectives at unmeasured well locations.

However, while acknowledging its limitations, Bayesian kriging remains potentially useful as a tool for summarizing large scale trends in arsenic concentrations or PC scores. Though it appears to underestimate high values, Bayesian kriging may still be useful at identifying regions where relatively high concentrations of arsenic or other groundwater constituents are more likely to be found. Therefore, these tools are potentially useful in estimating relative levels of exposure in human health risk studies. As a method that is accessible in commercial software packages, Bayesian kriging represents a flexible and viable statistical technique that could be used inform targeted well testing campaigns and potentially applied in smaller geographic areas where geostatistical analysis at smaller scales could be more informative.



## 4.7 References

- Ayotte, J.D., Nolan, B.T., Nuckols, J.R., Cantor, K.P., Robinson, Baris, D., Hayes, L., Karagas, M., Bress, W., Silverman, D.T., Lubin, J.H., 2006. Modeling the Probability of Arsenic in Groundwater in New England as a Tool for Exposure Assessment. *Environ. Sci. Technol.* 40, 3578–3585. doi:10.1021/es051972f
- Belkhiri, L., Boudoukha, A., Mouni, L., 2011. A multivariate statistical analysis of groundwater chemistry data. *Int. J. Environ. Res.* 5, 537–544.
- Canadian Association for Laboratory Accreditation Inc., n.d. CALA Directory of Laboratories. URL <http://www.caladirectory.ca/index.php>.
- Celik, I., Gallicchio, L., Boyd, K., Lam, T.K., Matanoski, G., Tao, X., Shiels, M., Hammond, E., Chen, L., Robinson, K.A., others, 2008. Arsenic in drinking water and lung cancer: a systematic review. *Environ. Res.* 108, 48–55.
- Chapagain, S.K., Pandey, V.P., Shrestha, S., Nakamura, T., Kazama, F., 2010. Assessment of deep groundwater quality in Kathmandu Valley using multivariate statistical techniques. *Water. Air. Soil Pollut.* 210, 277–288.
- Charrois, J.W.A., 2010. Private Drinking Water Supplies: Challenges for Public Health. *Can. Med. Assoc. J.* 182, 1061–1064. doi:10.1503/cmaj.090956
- Chen, C.-J., Wang, S.-L., Chiou, J.-M., Tseng, C.-H., Chiou, H.-Y., Hsueh, Y.-M., Chen, S.-Y., Wu, M.-M., Lai, M.-S., 2007. Arsenic and diabetes and hypertension in human populations: A review. *Toxicol. Appl. Pharmacol.* 222, 298–304. doi:10.1016/j.taap.2006.12.032
- Chen, Y., Graziano, J.H., Parvez, F., Liu, M., Slavkovich, V., Kalra, T., Argos, M., Islam, T., Ahmed, A., Rakibuz-Zaman, M., others, 2011. Arsenic exposure from drinking water and mortality from cardiovascular disease in Bangladesh: prospective cohort study. *Bmj* 342, d2431.
- Corkal, D., Schutzman, W.C., Hilliard, C.R., 2004. Rural water safety from the source to the on-farm tap. *J. Toxicol. Environ. Health A* 67, 1619–1642.
- Dental Health Promotion Working Group of Saskatchewan, 2011. Saskatchewan Community Fluoride Data 2010.
- Dunn, G., Bakker, K., Harris, L., 2014. Drinking Water Quality Guidelines across Canadian Provinces and Territories: Jurisdictional Variation in the Context of Decentralized Water Governance. *Int J Environ Res Public Health* 11, 4634–4651. doi:10.3390/ijerph110504634
- Environment and Climate Change Canada, 2007. Environment and Climate Change Canada - Water - Groundwater. URL <https://www.ec.gc.ca/eau-water/default.asp?lang=En&n=300688DC-1#sub5>.
- Farnham, I.M., Singh, A.K., Stetzenbach, K.J., Johannesson, K.H., 2002. Treatment of nondetects in multivariate analysis of groundwater geochemistry data. *Chemom. Intell. Lab. Syst.*, Fourth International Conference on Environ metrics and Chemometrics held

- in Las Vegas, NV, USA, 18-20 September 2000 60, 265–281. doi:10.1016/S0169-7439(01)00201-5
- Gong, G., Mattevada, S., O'Bryant, S.E., 2014. Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environ. Res.* 130, 59–69.
- Goovaerts, P., AvRuskin, G., Meliker, J., Slotnick, M., Jacquez, G., Nriagu, J., 2005. Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. *Water Resour. Res.* 41, W07013. doi:10.1029/2004WR003705
- Government of Saskatchewan, 2009. Sask H2O - Water Information Quick Facts. Sask H2O - Water Inf. Quick Facts. URL <http://www.saskh2o.ca/WaterInformation/QuickFacts.asp>.
- Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. *R Journal* 8, 204–218.
- Hassan, M.M., Atkins, P.J., 2011. Application of geostatistics with indicator kriging for analyzing spatial variability of groundwater arsenic concentrations in Southwest Bangladesh. *J. Environ. Sci. Health Part A* 46, 1185–1196.
- Health Canada, 2014. Guidelines for Canadian Drinking Water Quality - Summary Table. URL [http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/sum\\_guide-res\\_recom/index-eng.php](http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/sum_guide-res_recom/index-eng.php).
- Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M., Fernandez, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* 34, 807–816. doi:10.1016/S0043-1354(99)00225-0
- ISC, n.d. ISC - The Township System. URL <https://www.isc.ca/About/History/LandSurveys/MeasuringLandSask/Pages/TownshipSystem.aspx>.
- James, K.A., Meliker, J.R., Battenfield, B.E., Byers, T., Zerbe, G.O., Hokanson, J.E., Marshall, J.A., 2014. Predicting arsenic concentrations in groundwater of San Luis Valley, Colorado: implications for individual-level lifetime exposure assessment. *Environ. Geochem. Health* 36, 773–782.
- Kaiser, H.F., Rice, J., 1974. Little Jiffy, Mark Iv. *Educational and Psychological Measurement* 34, 111–117. doi:10.1177/001316447403400115
- Kim, D., Miranda, M.L., Tootoo, J., Bradley, P., Gelfand, A.E., 2011. Spatial Modeling for Groundwater Arsenic Levels in North Carolina. *Environ. Sci. Technol.* 45, 4824–4831. doi:10.1021/es103336s
- Krivoruchko, K., 2012. Empirical bayesian kriging. Esri Redlands CA USA. URL <http://www.esri.com/news/arcuser/1012/empirical-byesian-kriging.html>
- Lin, L.I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45, 255–268.
- Lisabeth, L.D., Ahn, H.J., Chen, J.J., Sealy-Jefferson, S., Burke, J.F., Meliker, J.R., 2010. Arsenic in Drinking Water and Stroke Hospitalizations in Michigan. *Stroke* 41, 2499–2504. doi:10.1161/STROKEAHA.110.585281

- Liu, C.-W., Lin, K.-H., Kuo, Y.-M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci. Total Environ.* 313, 77–89. doi:10.1016/S0048-9697(02)00683-6
- McLeod, L., Bharadwaj, L., Waldner, C., 2015. Risk factors associated with perceptions of drinking water quality in rural Saskatchewan. *Can. Water Resour. J. Rev. Can. Ressour. Hydr.* 40, 36–46. doi:10.1080/07011784.2014.985513
- Meliker, J.R., AvRuskin, G.A., Slotnick, M.J., Goovaerts, P., Schottenfeld, D., Jacquez, G.M., Nriagu, J.O., 2008. Validity of spatial models of arsenic concentrations in private well water. *Environ. Res.* 106, 42–50. doi:10.1016/j.envres.2007.09.001
- Moon, K., Guallar, E., Navas-Acien, A., 2012. Arsenic Exposure and Cardiovascular Disease: An Updated Systematic Review. *Curr. Atheroscler. Rep.* 14, 542–555. doi:10.1007/s11883-012-0280-x
- Navas-Acien, A., Silbergeld, E.K., Streeter, R.A., Clark, J.M., Burke, T.A., Guallar, E., 2005. Arsenic Exposure and Type 2 Diabetes: A Systematic Review of the Experimental and Epidemiologic Evidence. *Environ. Health Perspect.* 114, 641–648. doi:10.1289/ehp.8551
- Nazzal, Y., Zaidi, F.K., Ahmed, I., Ghrefat, H., Naeem, M., Al-Arifi, N.S.N., Al-Shaltoni, S.A., Al-Kahtany, K.M., 2015. The combination of principal component analysis and geostatistics as a technique in assessment of groundwater hydrochemistry in arid environment. *Curr. Sci.* 00113891 108, 1138–1145.
- Olea, R.A., 2006. A six-step practical approach to semivariogram modeling. *Stoch. Environ. Res. Risk Assess.* 20, 307–318.
- Onufrak, S.J., Park, S., Sharkey, J.R., Sherry, B., 2014. The relationship of perceptions of tap water safety with intake of sugar-sweetened beverages and plain water among US adults. *Public Health Nutr.* 17, 179–185. doi:10.1017/S1368980012004600
- Sánchez-Martos, F., Jiménez-Espinosa, R., Pulido-Bosch, A., 2001. Mapping groundwater quality variables using PCA and geostatistics: a case study of Bajo Andarax, southeastern Spain. *Hydrol. Sci. J.* 46, 227–242. doi:10.1080/02626660109492818
- Satyaji Rao, Y.R., Keshari, A.K., Gosain, A.K., 2009. Evaluation of regional groundwater quality using PCA and geostatistics in the urban coastal aquifer, East Coast of India. *Int. J. Environ. Waste Manag.* 5, 163–180.
- Shyu, G.-S., Cheng, B.-Y., Chiang, C.-T., Yao, P.-H., Chang, T.-K., 2011. Applying Factor Analysis Combined with Kriging and Information Entropy Theory for Mapping and Evaluating the Stability of Groundwater Quality Variation in Taiwan. *Int. J. Environ. Res. Public Health* 8, 1084–1109. doi:10.3390/ijerph8041084
- Sketchell, J., Shaheen, N., 2000. Ground water quality in rural Saskatchewan—Emerging issues for drinking water, in: *Maintaining Drinking Water Quality—Lessons from the Prairies and Beyond. Proceedings of the 9th National Conference on Drinking Water.* Regina, Saskatchewan, Canada. (Ed. W. Robertson.). pp. 242–258.

- Thompson, T.S., 2003. General Chemical Water Quality of Private Groundwater Supplies in Saskatchewan, Canada. *Bull. Environ. Contam. Toxicol.* 70, 0447–0454.  
doi:10.1007/s00128-003-0007-3
- Thompson, T.S., 2001. Nitrate Concentrations in Private Rural Drinking Water Supplies in Saskatchewan, Canada. *Bull. Environ. Contam. Toxicol.* 66, 64–70.  
doi:10.1007/s0012800206
- Thompson, T.S., Le, M.D., Kasick, A.R., Macaulay, T.J., 1999. Arsenic in Well Water Supplies in Saskatchewan. *Bull. Environ. Contam. Toxicol.* 63, 478–483.  
doi:10.1007/s001289901005
- Van Geen, A., Zheng, Y., Versteeg, R., Stute, M., Horneman, A., Dhar, R., Steckler, M., Gelman, A., Small, C., Ahsan, H., others, 2003. Spatial variability of arsenic in 6000 tube wells in a 25 km<sup>2</sup> area of Bangladesh. *Water Resour. Res.* 39, 1140.
- Water Security Agency, 2016. Municipal Drinking Water Quality Monitoring Guidelines. URL <http://www.saskh2o.ca/pdf/epb202.pdf>
- Water Security Agency, n.d. Saskatchewan's Drinking Water Quality Standards and Objectives (Summarized). URL <http://www.saskh2o.ca/pdf/epb507.pdf>
- Yang, Q., Jung, H.B., Culbertson, C.W., Marvinney, R.G., Loiselle, M.C., Locke, D.B., Cheek, H., Thibodeau, H., Zheng, Y., 2009. Spatial Pattern of Groundwater Arsenic Occurrence and Association with Bedrock Geology in Greater Augusta, Maine, USA. *Environ. Sci. Technol.* 43, 2714–2719.
- Yu, W.H., Harvey, C.M., Harvey, C.F., 2003. Arsenic in groundwater in Bangladesh: A geostatistical and epidemiological framework for evaluating health effects and potential remedies. *Water Resour. Res.* 39, 1146. doi:10.1029/2002WR001327

## **CHAPTER 5: ECOLOGICAL ANALYSIS OF ASSOCIATIONS BETWEEN GROUNDWATER QUALITY AND TYPE 2 DIABETES INCIDENCE AND PREVALENCE IN RURAL SASKATCHEWAN USING BAYESIAN HIERARCHICAL MODELS**

**Disclaimer:** this study is based in part on de-identified data provided by the Saskatchewan Ministry of Health. The interpretation and conclusions contained herein do not necessarily represent those of the Government of Saskatchewan or the Ministry of Health.

*The research uses the summarized water quality data from the previous chapter in an ecological analysis to investigate relationships between water quality and the occurrence of diabetes in rural Saskatchewan residents. Water quality risk factors derived from kriging of arsenic concentrations and principal component scores were used as exposure measures in Bayesian hierarchical models with the outcome of interest being incident and prevalent counts of diabetes cases derived from administrative health data. The investigation of associations between principal components summarizing health standards and aesthetic objectives and diabetes has not been previously reported. The principal component scores summarizing health standards reflect potential direct effects of water quality on the occurrence of diabetes, while the principal components summarizing aesthetic objectives represent indirect effects of poor water palatability.*

## 5.1 Abstract

With rates of diabetes increasing globally, there has been growing interest in the role of environmental exposures in the development of this disease. Arsenic in drinking water has been identified as a possible risk factor for diabetes in many parts of the world. Using existing administrative health and water quality surveillance data from rural Saskatchewan, associations were investigated between reported concentrations of arsenic, water health standards and aesthetic objectives and the incidence and prevalence of diabetes. Bayesian hierarchical models incorporating both spatial and unstructured random effects were compared to frequentist models with unstructured random effects. All models were adjusted for demographic and socio-economic factors as well as a surrogate measure for smoking rates.

Arsenic was not associated with an increased risk of diabetes incidence or prevalence. For private wells, having groundwater arsenic concentrations in the highest quintile was associated with decreased diabetes incidence in 2001-2012 (risk ratio=0.854, 95% credible interval 0.761-0.958) compared to the lowest quintile, a result inconsistent with other studies. This effect was not apparent in frequentist models for 2006-2009.

Having a score for the first principal component (PC) for health standards for public water supplies in the third quintile (risk ratio=1.101, 95% credible interval 1.019-1.188), fourth quintile (risk ratio=1.088, 95% credible interval 1.003-1.180), or fifth quintile (RR=1.115, 95% credible interval 1.026-1.213) was associated with an increase in diabetes prevalence compared to the first quintile. However, this result was not evident in any of the frequentist models from 2006-2009 and may have been a spurious finding. The first principal component for health standards in public water supplies primarily summarized selenium, nitrate and lead concentrations.

An increase in the principal component scores for the third aesthetic objective in private wells (characterized primarily by iron and manganese) was associated with decreased diabetes incidence. The association was apparent only in the Bayesian spatial model for 2010-2012, and a significant dose-response relationship was not evident. No other associations between the PC scores for either health standards or aesthetic objectives from public or private water supplies and diabetes were evident after accounting for spatial associations in the data. Based on this ecological analysis, there was no consistent evidence that water quality is associated with the occurrence of diabetes in residents of rural Saskatchewan.

## 5.2 Introduction

According to the Public Health Agency of Canada, the prevalence of diabetes in Canada increased by 70% between 1998/1999 and 2008/2009. Rates doubled in the 35-44 age group (Public Health Agency of Canada, 2011) mirroring steady increases in diabetes prevalence reported globally (World Health Organization, 2016). The province of Saskatchewan, Canada reported a 15% increase in diabetes prevalence between 2002/2003 and 2006/2007. An increased prevalence of diabetes is thought to be due primarily to an ageing population, decreasing mortality rates, and increasing rates of overweight and obesity in the population. However, interest in the potential contribution of environmental exposures to a range of chemicals and pollutants to increased rates of diabetes has grown in recent years (Prüss-Ustün et al., 2011; Thayer et al., 2012; Norman et al., 2013).

An increasing number of studies have identified associations between high concentrations of arsenic in drinking water and type 2 diabetes, although many questions remain about the dose-response relationship and biological mechanisms driving these associations (Maull et al., 2012; Kuo et al., 2013; Wang et al., 2014). Associations between low to moderate concentrations of arsenic in drinking water ( $<100 \mu\text{g/L}$ ) and diabetes are not as well established. Some studies have found positive associations (Navas-Acien et al., 2008; Del Razo et al., 2011; Bruner et al., 2014), while others have not demonstrated any links (Zierold et al., 2004; Steinmaus et al., 2009; Y. Chen et al., 2010; Li et al., 2013) between moderate concentrations of arsenic in drinking water and diabetes.

Unpalatable drinking water could also drive the consumption of beverages other than tap water including sugar-sweetened beverages (Onufrak et al., 2014). Because the consumption of sugar sweetened beverages has been linked to diabetes (Malik et al., 2010; de Koning et al., 2011), poor quality drinking water could potentially have indirect impacts on the development of type 2 diabetes.

It is estimated that approximately 15 % of Saskatchewan residents, primarily in rural areas, obtain their household water from private water sources (Government of Saskatchewan, 2002). As in many other jurisdictions, private water sources are not subject to regulation in Saskatchewan; owners bear full responsibility for testing their private supplies. In contrast, semi-public or public water supplies are subject to regulation under the Health Hazard Regulations or

Water regulations and are monitored by the Ministry of Health and the Water Security Agency, respectively. However, for public and semi-public supplies the requirements for testing water supplies vary according to the population served as well as the type of water source. Residents of smaller communities are therefore typically supplied with water that has not been tested with the same intensity as water supplied to larger population centers. This disparity in monitoring the safety of water supplies could potentially increase the risk of exposure to poor water quality for residents of rural areas.

In Canada, guidelines for drinking water quality and regulation are established by individual provinces, but are often based on the Canadian Guidelines for Drinking Water Quality. In Saskatchewan (SK), drinking water guidelines are categorized as standards, which are comprised of substances and organisms considered hazards to human health (e.g., heavy metals, microbes, pesticides), and objectives, which may impact the palatability of drinking water but are not considered health risks (Water Security Agency, n.d).

Previous studies have highlighted concerns about groundwater quality in Saskatchewan (Thompson et al., 1999; Sketchell and Shaheen, 2000; Thompson, 2001; Thompson, 2003). In particular, areas with moderately elevated arsenic concentrations in groundwater have been identified (Thompson et al., 1999). In addition, groundwater in Saskatchewan often has a high mineral content affecting its aesthetic qualities; high concentrations of iron, manganese, calcium, magnesium, and sulphate are common. One study reported that 99.6% of 535 tested wells exceeded at least one of Saskatchewan's Drinking Water Quality Standards and Objectives (Sketchell and Shaheen, 2000). A more recent survey of rural Saskatchewan residents found that 25% of respondents had complaints about the aesthetic quality of their household tap water (McLeod et al., 2015).

The primary hypothesis for this study was that residents in areas of rural Saskatchewan who are exposed to drinking water with higher concentrations of parameters, categorized as part of health standards or aesthetic objectives, are more likely to have higher rates of diabetes. Existing water quality and administrative health data provide an opportunity to examine this question using an ecological study design. While studies that can link exposure and disease in individuals provide stronger evidence for causal inference, they are costly and can be challenging when long-term exposures are difficult to quantify (Kunzli and Tager, 1997; Elliot and Savitz, 2008). Ecological



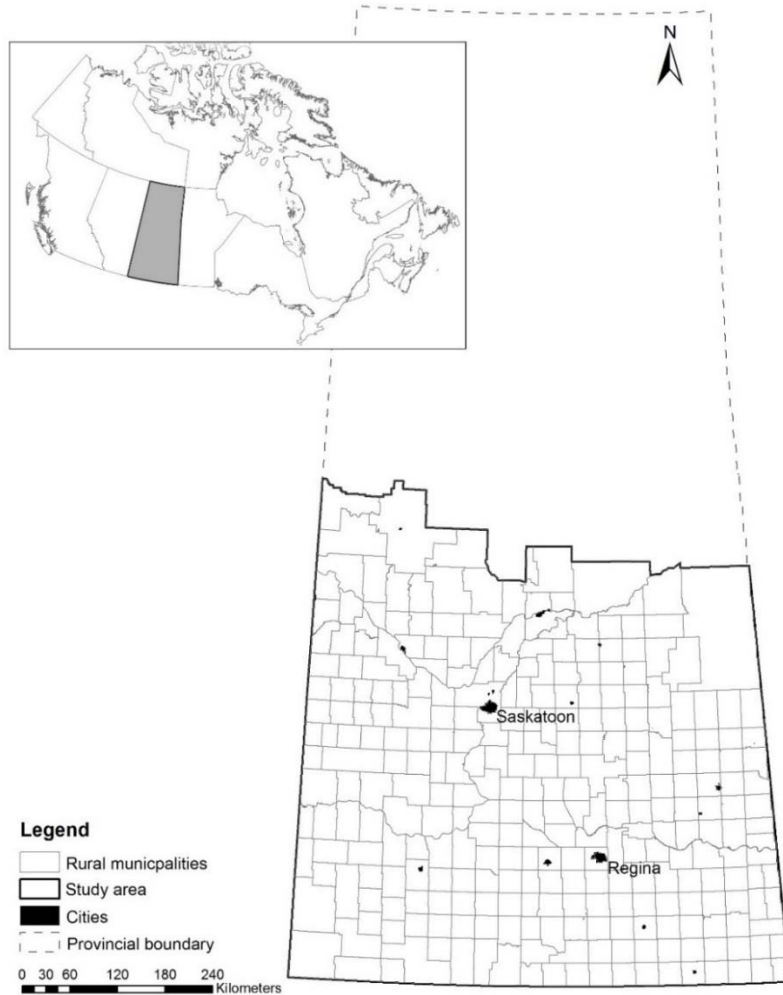
epidemiological studies must be interpreted cautiously, but can be useful for hypothesis screening (Rothman et al., 2008). It has been suggested that ecological studies are underused, especially when considering relatively rare outcomes and environmental exposures with long latent periods (de Vocht et al., 2013). The limitations of ecological studies can also be partially mitigated by using multi-level designs where important confounders are measured and linked to disease outcomes at the individual level (Kunzli and Tager, 1997; Elliot and Savitz, 2008). The primary goal of this study was to use existing surveillance data to investigate associations between groundwater quality and type 2 diabetes in rural Saskatchewan and to determine if more intensive studies of potential associations are warranted.

The first objective of this study was to use existing water surveillance data and population-based administrative health data to investigate associations between arsenic concentrations and type 2 diabetes in rural Saskatchewan. The second and third objectives of this study were to examine the associations between concentrations of chemicals in drinking water monitored either as part of health standards or as aesthetic objectives and the occurrence of type 2 diabetes in rural Saskatchewan.

## **5.3 Materials and Methods**

### *5.3.1 Study Area*

Because the population is very sparse in the northern part of Saskatchewan, restricting power, the study area was limited to the southern portion of the province (Figure 5.1). The northern border of the study area corresponds to the borders of the rural municipalities (RM), a local administrative unit in Saskatchewan.



**Figure 5.1** A map showing the province of Saskatchewan, the boundary of the study area for evaluation of associations between water quality and diabetes, and rural municipalities within the study area. Inset map shows the location of Saskatchewan within Canada.

### *5.3.2 Geographic Units for Analysis*

The geographic areas for use in the analysis were individual or aggregated RMs. To mitigate the problem of zero counts and very small populations at risk within each area after stratification by age, sex and First Nations status, RMs were aggregated with larger RMs when the population of residents over the age of 19 was less than 500. An algorithm for aggregating RMs was developed using the province's RM numbering system. RMs are numbered east to west and north to south; however, because RMs have been aggregated over time current RMs are not necessarily numbered consecutively with occasional large jumps in numbers between adjacent RMs. An RM

with a population of less than 500 was combined with the lowest-numbered adjacent RM, unless the RM numbers differed by more than two, in which case the RM was grouped with the highest-numbered adjacent RM.

In the administrative health data, individuals registered to a First Nation were historically assigned a residence code corresponding to the First Nation band affiliation rather than a geographic location. For residents with this type of residence code, the First Nation reserve with the greatest proportion of that First Nation's population in the 2006 Census of Canada was assumed to be the most likely place of residence. The Census Consolidated Subdivision (CCS) containing the most populous reserve for each First Nation was identified. Because CCS correspond geographically to RMs, the CCS number identified for each First Nation was used match each First Nation to an RM for aggregation. Cities, defined in the Saskatchewan Municipalities Act as a settlement with a population greater than 5000, were excluded from the geographic units used in the analysis.

Residential stability of the study participants was assessed by identifying the individuals who lived in the same RM in all years from 2004-2010 and reporting the proportion of all at-risk individuals who lived in the same area from 2004-2010.

### *5.3.3 Diabetes Data*

The population considered in developing the incidence and prevalence data used in this study included Saskatchewan residents age 35-74 with health care coverage as of June 30 in each year from 2002 through 2012 extracted from de-identified administrative health data.

Because Saskatchewan has universal health care coverage, all hospital and physician visits are billed to the province. Diabetes cases were identified according to the International Classification of Diseases (ICD) (Table 5.1) based on an algorithm previously validated in the literature (Hux et al., 2002; G. Chen et al., 2010; Dyck et al., 2010) and used by the Canadian Chronic Disease Surveillance System (Public Health Agency of Canada, 2009). A diabetes case was defined as an individual with one hospital visit or two physician visits within a two-year period billed with the relevant ICD codes; with the index date being assigned to the date of the first instance of a diabetes related ICD code defining that case. Cases of diabetes that could be related to pregnancy were excluded according to the criteria in Table 5.1.

**Table 5.1** Case definitions and International Classification of Diseases (ICD) codes used to identify cases of diabetes in Saskatchewan administrative health data.

Case Definition	ICD codes	Exclusions
2 physician claims within 2 years <i>Or</i> 1 hospital discharge abstract record	<u>ICD-9</u> <sup>1</sup> : 250 <u>ICD-10</u> <sup>2</sup> : E10-E14	Gestational diabetes: exclude any instances of diabetes ICD code with any obstetrical code occurring 120 days before or 180 days after: <u>ICD-9</u> : 641-676, V27 <u>ICD-10</u> : O1, O21-95, O98, O99, Z37

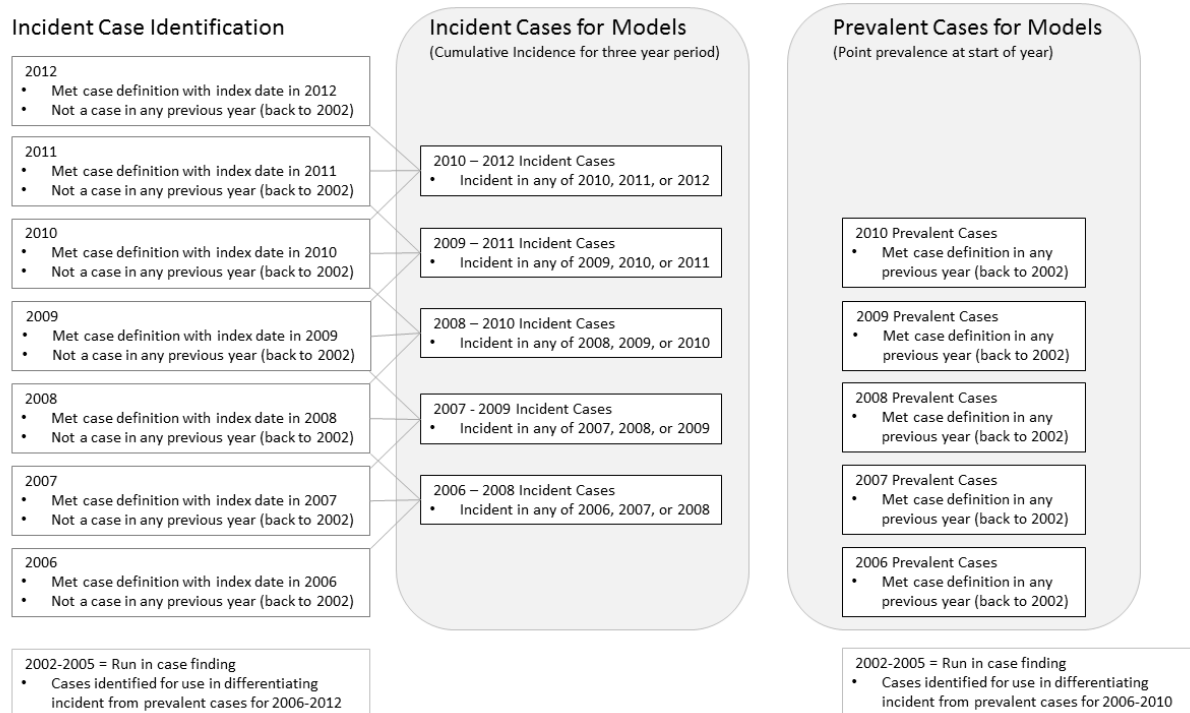
*Incident Case*: case definition met and not identified as case in previous year (back to 2002)

*Prevalent Case*: identified as a case in a previous year (back to 2002)

<sup>1</sup>International Classification of Diseases 9<sup>th</sup> Edition used to identify cases in Physician Services Claims File: Medical Services Branch.

<sup>2</sup>International Classification of Diseases 10<sup>th</sup> Edition used to identify cases in Hospital Discharge Abstract Database.

The number of incident cases of diabetes was aggregated over rolling three-year time periods for analysis to minimize the instances of zero counts. Five three-year periods were available for analysis from 2006-2008 through 2010-2012. Individuals who were identified as an incident case in any of the 3 years of a three-year period were considered an incident case for that period (Figure 5.2). Prevalent cases were analyzed for 5 one-year periods starting in 2006 and ending in 2010 for comparison to the incidence analysis, based on the point prevalence at the beginning of each year (Figure 5.2).



**Figure 5.2** Flow chart showing identification and aggregation of incident cases for the cumulative three-year incidence case counts used for modeling, and identification of prevalent cases.

Geographic location was assigned based on residence code in the first year of the three-year period to ensure that the corresponding water exposure preceded the time period for which cases were identified. Therefore, only individuals for which residence information was available in the first year of the three-year period were retained in the cohort. The age category was assigned based on age in the first year of the three-year period. Counts of incident cases of diabetes for each geographic area were stratified by sex, age category (34-44, 45-54, 55-64 and 65-74) and First Nations status. Ethics approval was obtained from the University of Saskatchewan Behavioral Research Ethics Board (Bio 12-332).

### 5.3.4 Water Data

Water quality surveillance data were accessed from the Water Security Agency for both public water supplies and private wells. Public water supply data were obtained from 1985-2012 and consisted of repeated samples from each supply taken as part of government-regulated monitoring of these supplies. The private well data were de-identified to protect well owners'

privacy; wells were sampled from 1996 to 2011 as part of a provincial Rural Water Quality Advisory Program which provided advice to owners of private wells as well as subsidized water testing available on a voluntary basis.

Trace metals and major ions that are included in the Saskatchewan Drinking Water Quality Standards and Objectives (Water Security Agency, n.d.) and that were routinely sampled in both the public supplies and private wells were included in the analysis. Arsenic, barium, boron, lead, nitrate, selenium and uranium were included in the analysis for health standards. Alkalinity, chloride, copper, hardness, iron, magnesium, manganese, sodium, sulphate, total dissolved solids, and zinc were included in the analysis for aesthetic objectives.

Where concentrations of sampled parameters were below detection limits, concentrations were imputed as  $\frac{1}{2}$  the detection limit of the method in use at the time of sampling. All water concentration data were right skewed and therefore log transformed to approximate a normal distribution before analysis.

Data from public water supplies were analyzed separately from data from private wells. The methods used to summarize exposure to the water quality variables are described in detail elsewhere (Chapter 4) and briefly summarized here. The geostatistical analysis to interpolate between data collection sites required a single value at each location. Therefore, generalized linear mixed models, including random effects for supply and a structured error term based on time between samples to account for the repeated measures, was used to estimate a single value for each parameter of interest for each public water supply using Proc Mixed in SAS (SAS Institute Inc., Cary, NC, USA).

The private well had only one measure per location. However, because the private well locations had been generalized to the centroid of the section of land (a parcel approximately 1.6 km x 1.6 km) on which they were situated to protect the well owners' privacy, private well locations had to be manually separated by adding or subtracting increments of 10 m to the latitude and longitude until no duplicate locations remained.

Arsenic concentrations were evaluated independently and as part of the health standards group. Groups of standards and objectives for each type of supply were summarized using principal components analysis (PCA) (SAS Institute Inc., Cary, NC, USA). Components with eigenvalues

greater than one were retained, then subject to varimax rotation to maximize the amount of variability explained by each principal component (PC). The resultant coefficients were used to calculate principal component scores for each retained component for water supply.

Empirical Bayesian kriging was used to interpolate values across the study area for arsenic (as log of the arsenic concentration in mg/L) and each of the principal factor scores which were summarized in rasters with a grid size of 800m x 800m for each variable (ArcGIS, ESRI, Redlands, CA). The mean values of the logged arsenic concentration and each principal component score for each type of supply were extracted for each geographic area for use in the epidemiological analysis (ArcGIS, ESRI, Redlands, CA) and the arsenic concentrations were back transformed and converted to µg/L for the exposure outcome analysis. The methods used to summarize the water data by PCA and kriging are described in detail elsewhere (Chapter 4).

### 5.3.5 Covariates

The stratum specific prevalence of chronic obstructive pulmonary disease (COPD) at the beginning of each time period analyzed was used as a proxy for smoking rate as a covariate in the final models. COPD cases were identified from administrative data for each year from 2002 to 2012 using ICD codes according to a published algorithm (Table 5.2) (Gershon et al., 2009). COPD prevalence, stratified by sex, age category and First Nations status, was calculated for each geographic unit for 2006-2010.

**Table 5.2** Case definitions and International Classification of Diseases codes used to identify cases of Chronic Obstructive Pulmonary Disease in Saskatchewan administrative health data.

Chronic Obstructive Pulmonary Disease	
≥ 1 physician claims within 1 year Or	<u>ICD-9<sup>1</sup></u> : 491, 492, 496
≥ 1 hospital discharge abstract	<u>ICD-10<sup>2</sup></u> : J41, J42, J43, J44
<i>Prevalent Case:</i> case definition met and/or identified as a case in a previous year (back to 2002)	

<sup>1</sup>International Classification of Diseases 9th Edition used to identify cases in Physician Services Claims File: Medical Services Branch.

<sup>2</sup>International Classification of Diseases 10<sup>th</sup> Edition used to identify cases in Hospital Discharge Abstract Database.

Education and total income data were obtained from the 2006 Census of Canada (Statistics Canada, 2014). First Nations status was not included in the education or income data from Statistics Canada, precluding stratification of these covariates by First Nations status. The proportion of residents not completing high school, stratified by sex and age category, was calculated for each area. To protect confidentiality, publically released census data for the population not completing high school was rounded to the nearest 5 and suppressed for areas with a population < 40 (Statistics Canada, 2008a). This meant the calculation of proportion not completing high school could be greater than 1. For age and sex categories within regions where the proportion not completing high school >1 or where the population at risk=0, the proportion not completing high school for that sex and age category for all Saskatchewan residents not residing in cities was imputed.

Average total annual income, stratified by sex and age, was calculated for each region. However, the age categorization available for the income data was not an exact match for the age categorization used for the disease data. Therefore, the average total annual income for those aged 25-44 from the Census data was used for the 34-44 age category, income for those aged 45-64 from the Census data was used for both the 45-54 and 55-64 age categories, and income for those 65 and over in the Census data was assigned to the 65-74 age group. For confidentiality, income data is suppressed for populations less than 250 (Statistics Canada, 2008b); for regions where the age and sex categories were missing income data, the average total annual income for that age and sex category for Saskatchewan residents not residing in cities was imputed. Average total annual income was centered on the mean for Saskatchewan residents in the corresponding age and sex category and not residing in cities and scaled so that regression coefficients represented a change in diabetes incidence or prevalence for each \$1000 of personal income.

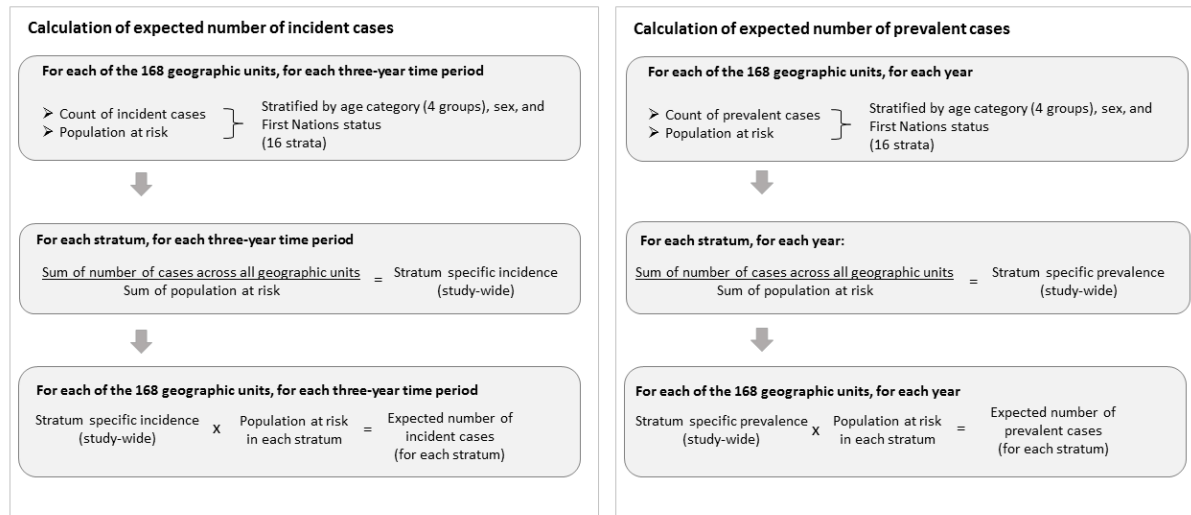
### *5.3.6 Bayesian Statistical Analyses of Associations between Water Quality Data and Diabetes*

Generalized linear mixed models with a log link function were used to investigate associations between water quality variables and incident and prevalent cases of diabetes. Counts of cases were assumed to follow a Poisson distribution with mean  $\mu_i = E_i \theta_i$  where  $E_i$  is the expected count, and  $\theta_i$  is the standardized morbidity ratio (SMR).

Stratum-specific expected case counts for diabetes incidence and prevalence were calculated for each area for each time period during the study. To obtain the numerator of the expected



incidence or prevalence for each stratum, or combination of sex, age category, and First Nations status, the total number of cases for that time period was summed for all regions in the study area. Similarly, to obtain the denominator for each stratum, the total number of people at risk within each stratum was summed across all geographic regions in the study area for each of the time periods. The average incidence and prevalence for each stratum for each time period were calculated by dividing the total number of cases for each combination of sex, age and First Nations status by the corresponding population at risk. These stratum-specific study-wide expected incidence and prevalence rates were then multiplied by the population at risk for each sex, age, and First Nations stratum for each of the geographic areas to obtain area- and stratum-specific expected counts of incident and prevalent cases (Figure 5.3).



**Figure 5.3** Summary of the process and equations used to obtain stratum-specific expected case counts for diabetes incidence and prevalence for each geographic unit.

The 2010-2012 incidence and 2010 prevalence data were analyzed using Bayesian models based on the hierarchical model proposed by Besag, York and Mollie (1991), which incorporates spatially correlated random effects as well as unstructured random effects (equation 5.1). The spatially structured random effect ( $v_i$ ) was modeled using an intrinsic conditionally autoregressive prior distribution where each random effect was assumed to follow a normal distribution whose mean and precision depend on neighboring random effects, where  $\delta_i$  is the set of adjacent neighbors for the  $i$ th geographic area and  $n_{\delta_i}$  is the number of neighbors for area  $i$ . This framework smooths estimates by borrowing information from neighboring areas (Lawson,

2013). The unstructured random effects ( $h_i$ ) were modeled using a zero mean Gaussian distribution. Hyperpriors for the precision on the structured ( $\tau_v$ ) and unstructured ( $\tau_h$ ) random effects were given gamma (0.001, 0.001) distributions. The regression coefficients ( $\beta_{1...k}$ ) for each of  $k$  fixed effects ( $x_{1...k}$ ) and for the intercept ( $\beta_0$ ) were assumed to follow a zero mean Gaussian distribution and were given uninformative priors.

$$\log \mu_i = \log E_i + \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + h_i + v_i \quad (5.1)$$

$$\beta_0, \beta_{1...k} \sim \text{Normal}(0, 10^{-5})$$

$$h_i \sim \text{Normal}(0, \tau_h)$$

$$v_i | \mathbf{v}_{-i} \sim \text{Normal}(\bar{v}_{\delta_i}, \tau_v / n_{\delta_i})$$

$$\bar{v}_{\delta_i} = \frac{1}{n_{\delta_i}} \sum_{j \in \delta_i} v_j$$

$$\tau_h, \tau_v \sim \text{gamma}(0.001, 0.001)$$

OpenBUGS 3.2.3 (Lunn et al., 2009) was used for Bayesian modeling and to generate the adjacency matrix identifying neighbors based on queen contiguity for the geographic areas for use in spatial analysis.

Bayesian models were built separately for arsenic concentrations, aesthetic objective factor scores, and the health standards factor scores, for public and private water supplies, for a total of six models for diabetes incidence (2010-2012 accumulated incidence) and six models for diabetes prevalence (2010). COPD prevalence (stratified by age, sex and First Nations status), education level summarized as the proportion not completing high school (stratified by age and sex), and average total annual income in thousands of dollars (stratified by age and sex) were included as covariates in each model. Water quality risk factors were first evaluated as continuous variables in the models. Where potential associations between water variables and diabetes incidence or prevalence were identified, the linearity assumption was examined by adding a quadratic term for the water variable to the model and assessing if the credible interval

for the squared term included the null value of 1. Furthermore, variables for which potential associations were identified were categorized into quintiles with the lowest quintile (Q1) set as the reference category. The relationship between the categorized coefficients was examined to further characterize whether the form of the relationship between the risk factor and outcome was monotonic.

For each model, three chains were initiated and parameters were monitored until convergence was reached according to the Brooks-Gelman-Rubin diagnostic (Gelman and Rubin, 1992). Convergence was assessed both visually using plots and quantitatively using the CODA (Plummer et al., 2006) package in R 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria), using the criteria that the 97.5% upper bound of the scale reduction factor was 1.05 or less. After the burn-in period, 20,000 iterations from each of three chains for a total of 60,000 iterations were sampled to get estimates for the model parameters. Adequate sampling was confirmed by ensuring the MCMC error was less than 5% of the sample standard deviation for each monitored parameter.

*5.3.7 Frequentist Statistical Analyses of Associations between Water Quality Data and Diabetes*  
Data from each of the five three-year time periods for incidence and each year from 2006-2010 for prevalence were also analyzed as a frequentist generalized linear mixed model GLMM using Proc Glimmix in SAS (SAS Institute Inc., Cary, NC, USA) using a log link and a Laplace approximation, with a single random effect for region. The results were compared with the Bayesian models for incidence in 2010-2012 and prevalence in 2010 to assess consistency in the results across time periods. When the Bayesian model had not identified an association, significant results were targeted for discussion from the frequentist models only where the result was significant in at least 4 of the 5 time periods examined.

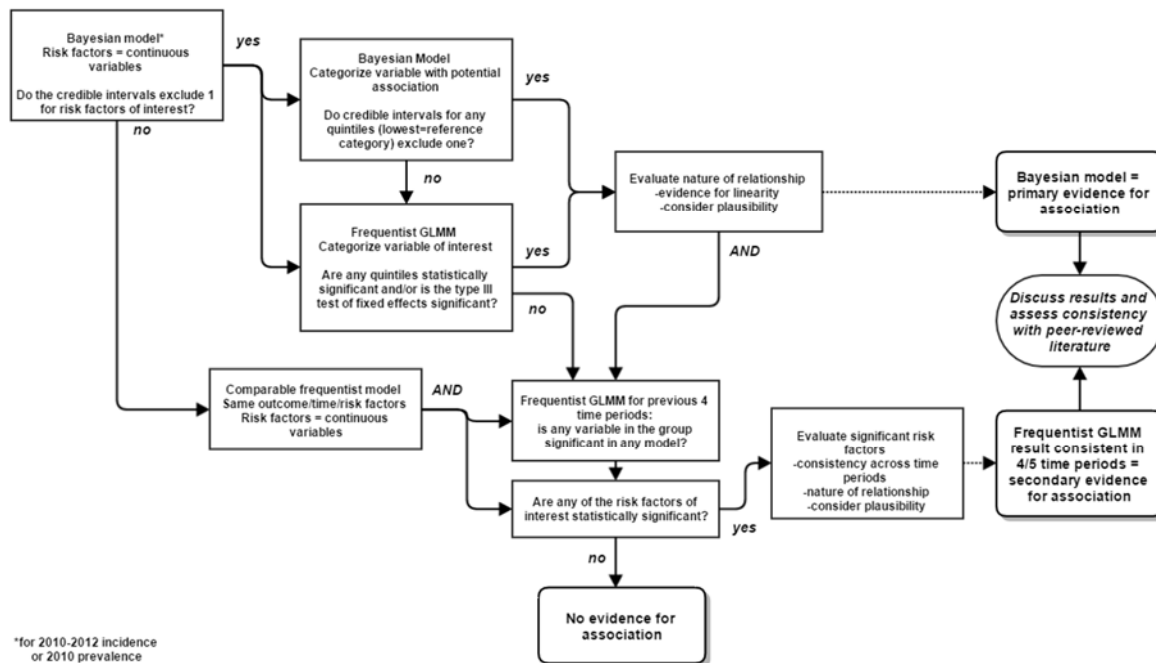
#### *5.3.8 Model Results and Assessment of Fit*

Regression coefficients were exponentiated to produce a risk ratio (RR), which represented the multiplicative effect of risk factors on the SMR. RRs were reported as point estimates along with 95% credible intervals (CrI) for Bayesian models and 95% confidence intervals (CI) for frequentist models.

The random effect residuals from the frequentist models for 2010-2012 for diabetes incidence and for 2010 for diabetes prevalence were examined for spatial clustering by calculating Moran's I in GeoDA 1.8.10 (The Center for Spatial Analysis, Chicago, Illinois) based on a spatial structure using first order queen contiguity.

To assess whether model fit was improved by inclusion of the spatial random effects in addition to unstructured random effects, the deviance information criterion (DIC), a measure of model fit, was compared for models with and without spatially structured random effects (Spiegelhalter et al., 2002). Comparisons were made for models where a 95% CrI of the effect estimate for the water quality variable of interest did not include 1. To ensure valid comparisons between DIC values, the value of pD, a measure of model complexity representing the effective number of parameters was considered along with the DIC (Spiegelhalter et al., 2002). If  $pD < 0$ , the DIC is not interpretable.

To facilitate assessment of potential associations between the drinking water risk factors and diabetes incidence or prevalence, a decision tree approach was taken (Figure 5.4). To identify associations for further discussion, results were examined from the Bayesian and frequentist models for 2010-2012 diabetes incidence and 2010 diabetes prevalence and the frequentist models for the preceding 4 time periods (Figure 5.4). A total of 15 risk factors were assessed for associations with diabetes incidence and prevalence for each time period: arsenic concentrations (one each for public water supplies and private wells), health standards PCs (four PCs from public water supplies and three from private wells), and aesthetic objectives PCs (three PCs each for public water supplies and private wells). A conservative experiment-wise error rate would therefore be based on 15 tests for each time period for each outcome. The probability of falsely identifying the observed numbers of associations given 15 tests was calculated using a modified hypergeometric distribution in publically available software (FreeCalc, EpiTools, AusVet Animal Health Services). For example, using this calculator the probability of having a single false positive is equal to an experiment-wise error  $= 1 - (1 - \alpha)^k$  for  $k$  tests. For this study, the probability of finding at least the observed number of positive results was estimated to assess the potential for experiment-wise type 1 error, if no true associations existed.



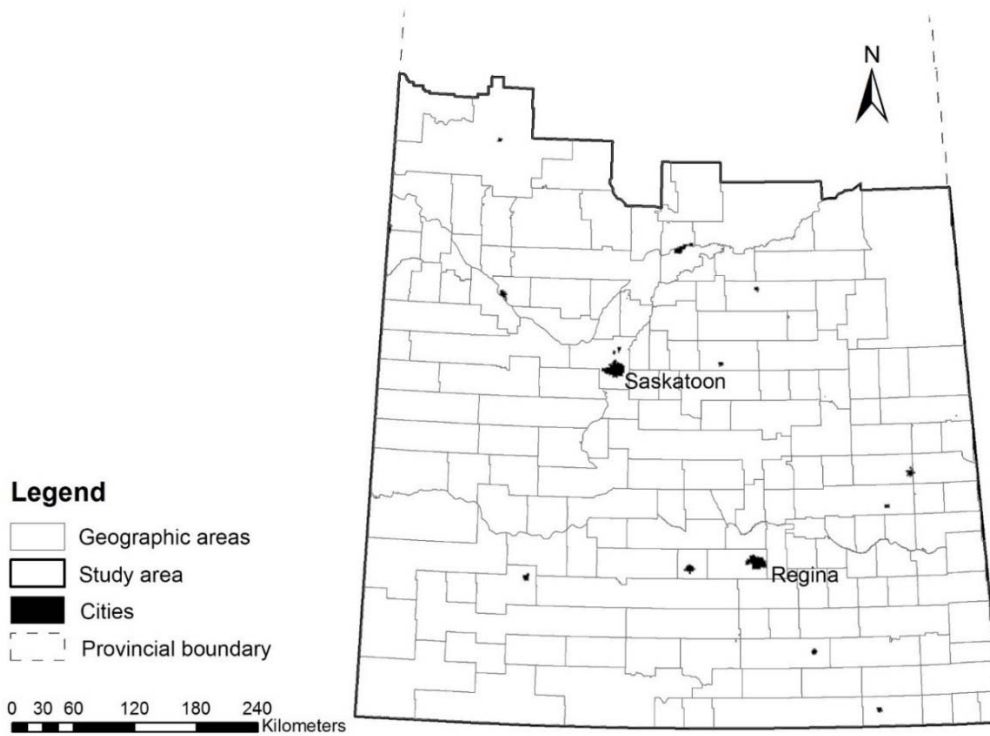
**Figure 5.4** A flow chart outlining process of evaluating potential associations between water quality risk factors and diabetes outcomes in Bayesian and frequentist models.

Sensitivity to the choice of hyperprior for the random effects was assessed for by comparing the effect of different hyperprior specifications for one of the significant associations. The value of effect estimates for arsenic concentration were compared in models with a  $\text{gamma}(0.5, 0.0005)$  placed on the precision for the random effects, and a  $\text{uniform}(0, 5)$  distribution placed on the standard deviation for the precision, where precision is equal to the inverse of the squared standard deviation (Jang et al., 2007; Lawson, 2013).

## 5.4 Results

Amalgamation of 296 RMs over the study area resulted in the delineation of 168 geographic areas for use in the analysis (Figure 5.5). Stratifying by sex, First Nations status and four age categories for 168 geographic areas results in 2,688 possible strata, with each stratum representing a row of data for the analysis. After excluding strata with population=0 from the analysis 2041, 2024, 2010, 2001, and 1982 observations were used for the diabetes incidence models for 2010-2012, 2009-2011, 2008-2010, 2007-2009, and 2006-2008 respectively, and

2114, 2090, 2085, 2064, and 2046 observations were available for the prevalence models for 2010, 2009, 2008, 2007, and 2006 respectively.



**Figure 5.5** A map of Saskatchewan showing geographic regions for analysis of the association between measures of water quality and diabetes developed by amalgamation of RMs.

The population of residents aged 35-74 in the study area for which residence information was available ranged from 178,324 to 188,423 for 2006-2010. The population in most age groups increased over the course of the study, with the exception of the 35-44 age group which decreased in size (Table 5.3). In 2010, the proportion of residents in rural Saskatchewan in the 35-44 age group was 23.4% (44,163/188,423), while there were 31.7% (59,718/188,423) in the 45-54 age group, 27.6% (51,932/188,423) in the 55-64 age group, and 17.3% (32,610/188,423) in the 65-74 age group. The percent of women in the study population was 48.7% (91,742/188,423) and the percent reported as First Nation was 7.3% (13,797/188,423). In 2010, the population at risk within geographic units ranged from 316 to 6,390 (median=913). Tracking the RM of residence for the study population from 2004-2010 revealed that 76.5% of residents lived in the same geographic area for the entire study period.

**Table 5.3** Number of rural Saskatchewan residents included in the study by year and by age group.

<b>Base Year<sup>1</sup></b>	<b>Total <i>n</i></b>	<b>By age group (n)</b>				<b>By sex (n)</b>		<b>First Nations (n)</b>
		<b>35-44</b>	<b>45-54</b>	<b>55-64</b>	<b>65-74</b>	<b>Male</b>	<b>Female</b>	
2006	178324	46495	57644	43269	30916	91304	87020	12048
2007	179496	45193	58496	44959	30848	92038	87458	12347
2008	183250	44851	59603	47404	31392	94035	89215	12810
2009	185154	44102	59599	49590	31863	94939	90215	13318
2010	188423	44163	59718	51932	32610	96681	91742	13797

<sup>1</sup> Because place of residence was determined based on the first year of the three-year period used to estimate cumulative incidence, the population in the first year of each three-year time period determined the population at-risk for the respective time period

In each consecutive three-year time period, the observed cumulative incidence of diabetes decreased, while the observed prevalence increased in consecutive years (Table 5.4). The median cumulative incidence for all residents aged 35-74 among geographic units for 2010-2012 was 2.8% (5<sup>th</sup> percentile = 1.7% and 95<sup>th</sup> percentile=4.2%). The observed incidence for 2010 was 2.3% for females and 3.4% in males and ranged from 1.4% for those aged 35-44 to 4.9% for those in the 65-74 age group. The median diabetes prevalence among geographic units in 2010 was 9.3% (5<sup>th</sup> percentile=6.5%, 95<sup>th</sup> percentile=14.7%). The observed prevalence for 2010 was 8.9% for females and 11.1% in males and ranged from 3.9% among 35-44 year olds to 20.2% for those in the 65-74 age group.

**Table 5.4** Diabetes cumulative incidence and prevalence in each time period analyzed for the population of the study area in southern Saskatchewan ages 35 to 75 who did not live in a city.

Total (n)	Incident Diabetes			Year	Prevalent Diabetes	
	3 Year Period	Incident cases (n)	Incidence (%)		Prevalent cases (n)	Prevalence (%)
178324	2006-2008	5585	3.4	2006	14346	8.0
179496	2007-2009	5426	3.3	2007	15556	8.7
183250	2008-2010	5352	3.2	2008	16705	9.1
185154	2009-2011	5311	3.2	2009	17771	9.6
188423	2010-2012	4815	2.8	2010	18871	10.0

The median and 95<sup>th</sup> percentile arsenic concentrations in the raw public water supply data were 0.9 µg/L and 14 µg/L respectively. After using a GLMM to summarize historical data for each location, the median predicted arsenic for public water supplies was 0.8 µg/L and the 95<sup>th</sup> percentile was 7.5 µg/L. For the observed private well data, the median arsenic concentration was 0.9 µg/L and the 95<sup>th</sup> percentile was 23 µg/L. For public supplies, 6.9 % of samples exceeded the Saskatchewan drinking water standard of 10 µg/L, while 13.5 % exceeded the standard for the private well samples. Whereas, 22.9 % of samples were below detection limits in the public supply data, and 21.3 % were below detection limits among the private well samples (Chapter 4).

For public water supplies, the predicted median of the arsenic concentrations summarized for each of the 168 geographic areas in the study was 0.9 µg/L, with a 95<sup>th</sup> percentile of 1.9 µg/L. For private water supplies, the predicted median was 1.3 µg/L and the 95<sup>th</sup> percentile was 3.1 µg/L.

The cut points defining the quintiles for arsenic concentrations for the public water supply data were 0.73 µg/L, 0.88 µg/L, 1.01 µg/L, and 1.31 µg/L. For the private well data, the cut points defining quintiles of arsenic concentration were 0.86 µg/L, 1.16 µg/L, 1.55 µg/L, and 2.25 µg/L.

PCA yielded three retained principal components (PC<sub>health</sub>) for health standards from both the public and private water supply data (Table 5.5); however, they differed somewhat in their loadings (Chapter 4).



PCA yielded four aesthetic objective principal components ( $PC_{\text{aesthetic}}$ ) for the public water supply data, and three for the private supply data (Table 5.5), but the loadings for the first two components were very similar (Chapter 4).

**Table 5.5** Summary of principal components analysis of groundwater parameters used as risk factors in models, showing component loadings along with the eigenvalue and cumulative variances for each retained component. The maximum loading for each parameter is indicated in bold. Details of the PCA analysis can be found in Chapter 4.

<i>Health Standards</i>	<b>Public Water Supplies</b>			<b>Private Wells</b>		
	$PC1_{\text{health}}$	$PC2_{\text{health}}$	$PC3_{\text{health}}$	$PC_{\text{health}}$	$PC2_{\text{health}}$	$PC3_{\text{health}}$
Arsenic	-0.121	0.142	<b>0.808</b>	-0.341	-0.091	<b>0.474</b>
Barium	0.047	<b>-0.818</b>	-0.141	-0.041	<b>0.893</b>	0.100
Boron	-0.062	<b>0.903</b>	-0.123	-0.195	<b>-0.818</b>	0.168
Lead	<b>0.472</b>	0.092	0.171	0.156	0.026	<b>0.893</b>
Nitrate	<b>0.768</b>	-0.071	-0.164	<b>0.770</b>	0.275	-0.110
Selenium	<b>0.867</b>	-0.220	0.019	<b>0.853</b>	-0.007	0.074
Uranium	0.387	-0.290	<b>0.576</b>	<b>0.772</b>	-0.013	-0.049
Eigenvalue	2.127	1.275	1.059	2.290	1.381	1.057
Cumulative variance (%)	30.4	48.6	63.7	32.7	52.5	67.6

<i>Aesthetic Objectives</i>	$PC1_{\text{aesthetic}}$	$PC2_{\text{aesthetic}}$	$PC3_{\text{aesthetic}}$	$PC4_{\text{aesthetic}}$	$PC1_{\text{aesthetic}}$	$PC2_{\text{aesthetic}}$	$PC3_{\text{aesthetic}}$
	$PC1_{\text{aesthetic}}$	$PC2_{\text{aesthetic}}$	$PC3_{\text{aesthetic}}$	$PC4_{\text{aesthetic}}$	$PC1_{\text{aesthetic}}$	$PC2_{\text{aesthetic}}$	$PC3_{\text{aesthetic}}$
Alkalinity	<b>0.755</b>	0.111	0.164	-0.198	<b>0.687</b>	-0.023	0.217
Chloride	<b>0.753</b>	-0.193	0.002	0.226	<b>0.779</b>	0.043	-0.195
Copper	0.127	0.012	-0.200	<b>0.714</b>	0.030	0.223	<b>-0.757</b>
Hardness	0.009	<b>0.973</b>	0.066	0.042	0.067	<b>0.960</b>	0.038
Iron	0.138	-0.089	<b>0.901</b>	0.053	0.117	0.121	<b>0.784</b>
Magnesium	-0.014	<b>0.961</b>	0.055	0.038	0.103	<b>0.951</b>	0.020
Manganese	0.188	0.452	<b>0.711</b>	-0.065	0.062	0.468	<b>0.663</b>
Sodium	<b>0.914</b>	-0.199	0.136	0.026	<b>0.922</b>	-0.116	0.118
Sulphate	<b>0.663</b>	0.517	0.018	0.116	<b>0.609</b>	0.555	0.076
Total Dissolved Solids	<b>0.920</b>	0.288	0.121	-0.016	<b>0.907</b>	0.325	0.082
Zinc	-0.089	0.078	0.237	<b>0.763</b>	-0.091	<b>0.396</b>	-0.375
Eigenvalue	3.746	2.362	1.264	1.181	3.775	2.184	1.779
Cumulative variance (%)	34.1	55.5	67.0	77.8	34.3	54.2	70.4

Similar to arsenic concentrations, PC scores predicted by kriging and averaged over the geographic regions exhibited less variability and a smaller range than the data for individual water supplies which were used as input for kriging (Table 5.6).

**Table 5.6** Median and 5<sup>th</sup> and 95<sup>th</sup> percentile values for the estimated mean exposures for principal component scores for each geographic and the corresponding values for individual water supplies used as input for kriging for public water supplies and private wells.

Variable	Mean area predicted values (after kriging)				PCA results (point values before kriging)			
	n	P5	Median	P95	n	P5	Median	P95
<b><i>Public water supplies</i></b>								
PC1 <sub>health</sub>	168	-0.55	0.09	0.52	459	-1.47	0.01	1.83
PC2 <sub>health</sub>	168	-0.66	0.04	0.69	459	-1.68	0.18	1.49
PC3 <sub>health</sub>	168	-0.62	0.05	0.68	459	-1.60	0.08	1.59
PC1 <sub>aesthetic</sub>	168	-0.63	0.02	0.68	435	-1.66	0.11	1.59
PC2 <sub>aesthetic</sub>	168	-0.78	0.07	0.65	435	-2.32	0.23	1.13
PC3 <sub>aesthetic</sub>	168	-0.56	-0.07	0.63	435	-1.74	0.08	1.55
PC4 <sub>aesthetic</sub>	168	-0.48	0.01	0.41	435	-1.56	-0.06	1.81
<b><i>Private wells</i></b>								
PC1 <sub>health</sub>	168	-0.55	0.05	0.63	3970	-1.51	-0.09	1.73
PC2 <sub>health</sub>	168	-0.70	-0.15	0.73	3970	-1.48	-0.09	1.70
PC3 <sub>health</sub>	168	-0.30	0.17	0.57	3970	-1.67	0.09	1.39
PC1 <sub>aesthetic</sub>	168	-0.69	0.12	0.78	3999	-1.72	0.06	1.53
PC2 <sub>aesthetic</sub>	168	-0.91	0.18	0.61	3999	-2.20	0.15	1.23
PC3 <sub>aesthetic</sub>	168	-0.41	0.07	0.50	3999	-1.73	0.09	1.49

P5 = 5<sup>th</sup> percentile, P95 = 95<sup>th</sup> percentile

Among geographic units, the median observed COPD prevalence in 2010 for all residents aged 35-74 was 5.5% (5<sup>th</sup> percentile=3.2% and 95<sup>th</sup> percentile=9.1%). From 2006-2010 the prevalence was lowest in 2006 with a median of 3.7% among geographic units, increasing each year through 2010.

Based on Census of Canada 2006 results, of the population age 35-74 in the study area not living in cities, the median proportion not completing high school was 29.6% (5<sup>th</sup> percentile = 17.0% and 95<sup>th</sup> percentile = 42.9%). The median average total personal income reported for residents in the study area not living in cities was \$27,375 (5<sup>th</sup> percentile = \$19,893 and 95<sup>th</sup> percentile = \$37,169).

#### *5.4.1 Associations between Arsenic Concentration and Diabetes Incidence*

There was no association demonstrated between groundwater arsenic concentration in public water supplies and diabetes incidence in the Bayesian or frequentist Poisson regression analysis for 2010-2012 (Table 5.7).

Initial analysis of arsenic in private wells as a continuous variable suggested that as groundwater arsenic concentration increased, the incidence of diabetes decreased (RR=0.954, 95% CrI 0.913-0.997, with a probability that  $RR < 1$  of 98.1%). However, categorization of the arsenic concentration into quintiles demonstrated that the relationship was not linear or monotonic (Table 5.7). The only comparison between exposure quintiles where the CrI did not include 1 was for the highest quintile of arsenic compared to the lowest (RR=0.854, 95% CrI 0.761-0.958).

When considering arsenic in private wells as both a continuous (RR=0.956, 95% CI 0.917-0.998,  $p=0.04$ ) and categorical variable (Table 5.7), the effect estimates for comparable frequentist GLMMs with single random effects for region were similar to those from the Bayesian analysis. The frequentist GLMM association between arsenic concentration in private wells and diabetes incidence was only significant when comparing the highest arsenic quintile to the lowest (RR=0.859, 95% CI 0.771-0.957); overall the categorized arsenic variable was significant according to the type III test of fixed effects ( $p=0.04$ ).

**Table 5.7** Comparison of Bayesian and frequentist models of the association between arsenic concentrations in drinking water and cumulative diabetes incidence 2010-2012 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public Water Supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.010	0.913	1.119	Intercept	1.006	0.912	1.110	0.90
Arsenic (µg/L)	0.980	0.910	1.054	Arsenic (µg/L)	0.983	0.918	1.053	0.62
COPD prevalence	1.830	1.070	3.096	COPD prevalence	1.853	1.087	3.159	0.02
Education <sup>1</sup>	0.886	0.741	1.056	Education <sup>1</sup>	0.886	0.743	1.056	0.18
Income <sup>2</sup>	0.998	0.994	1.001	Income <sup>2</sup>	0.998	0.995	1.001	0.23
Random effects		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.090	0.030	Unstructured RE		0.013	0.005	
Spatially structured RE		0.064	0.031					
Burn in period = 120,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.054	0.955	1.161	Intercept	1.054	0.958	1.159	0.28
Arsenic Quintile 1	reference category			Arsenic Quintile 1	reference category			0.04 <sup>3</sup>
Arsenic Quintile 2	0.933	0.833	1.046	Arsenic Quintile 2	0.931	0.832	1.041	0.21
Arsenic Quintile 3	0.999	0.895	1.119	Arsenic Quintile 3	0.987	0.887	1.098	0.81
Arsenic Quintile 4	0.937	0.840	1.047	Arsenic Quintile 4	0.938	0.845	1.042	0.23
Arsenic Quintile 5	<b>0.854</b>	<b>0.761</b>	<b>0.958</b>	Arsenic Quintile 5	<b>0.859</b>	<b>0.771</b>	<b>0.957</b>	<b>0.006</b>
COPD prevalence	1.816	1.055	3.086	COPD prevalence	1.863	1.095	3.169	0.02
Education <sup>1</sup>	0.879	0.736	1.049	Education <sup>1</sup>	0.880	0.739	1.049	0.16
Income <sup>2</sup>	0.998	0.995	1.001	Income <sup>2</sup>	0.998	0.995	1.001	0.24
Random effects		SD	SE	Random Effect		Variance	SE	
Unstructured RE		0.077	0.029	Unstructured RE		0.010	0.005	
Spatially structured RE		0.070	0.028					
Burn in period = 100,000 iterations								
Number of observations = 2041								

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars). <sup>3</sup>Overall p value for categorized variable based on type III likelihood ratio test.

Comparison of the effect estimates for arsenic concentration as a continuous variable in a frequentist GLMM for the 4 previous time periods indicated there were no significant associations between groundwater arsenic concentrations in public water supplies or private wells and diabetes incidence in these time periods (Table 5.8).

**Table 5.8** Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and 3-year cumulative incidence of diabetes for the time periods 2006-2008 through 2009-2011 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009-2011				2008-2010				2007-2009				2006-2008			
	RR	95% CI			RR	95% CI			RR	95% CI			RR	95% CI		
		lower	upper	p		lower	upper	p		lower	upper	p		lower	upper	p
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.982	0.898	1.074	0.70	0.974	0.890	1.067	0.57	0.987	0.899	1.084	0.79	0.983	0.891	1.085	0.73
Arsenic (µg/L)	1.0009	0.948	1.073	0.79	1.011	0.950	1.077	0.73	0.995	0.932	1.063	0.88	0.989	0.921	1.062	0.76
COPD prevalence	1.728	1.039	2.871	0.04	1.398	0.822	2.377	0.22	1.424	0.815	2.486	0.22	1.193	0.651	2.185	0.57
Education <sup>1</sup>	0.915	0.776	1.080	0.30	0.970	0.823	1.142	0.71	0.969	0.823	1.141	0.71	1.019	0.866	1.199	0.82
Income <sup>2</sup>	0.998	0.995	1.001	0.22	0.999	0.995	1.002	0.37	0.997	0.994	1.000	0.09	0.999	0.995	1.002	0.44
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.008		0.004		0.009		0.004		0.013		0.005		0.020		0.006	
<hr/>																
<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	RR	95% CI			RR	95% CI			RR	95% CI			RR	95% CI		
		lower	upper	p		lower	upper	p		lower	upper	p		lower	upper	p
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	1.051	0.965	1.144	0.26	1.036	0.950	1.129	0.43	1.033	0.945	1.128	0.47	1.011	0.921	1.109	0.82
Arsenic (µg/L)	0.964	0.927	1.001	0.06	0.969	0.932	1.008	0.12	0.968	0.930	1.009	0.13	0.975	0.933	1.019	0.26
COPD prevalence	1.712	1.031	2.842	0.04	1.396	0.822	2.371	0.22	1.413	0.809	2.466	0.23	1.187	0.648	2.173	0.58
Education <sup>1</sup>	0.918	0.778	1.083	0.31	0.971	0.824	1.143	0.72	0.970	0.824	1.143	0.72	1.020	0.867	1.200	0.81
Income <sup>2</sup>	0.998	0.995	1.001	0.22	0.999	0.995	1.002	0.37	0.997	0.994	1.000	0.10	0.999	0.995	1.002	0.44
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.007		0.004		0.009		0.004		0.012		0.005		0.020		0.006	
<hr/>																
n=2024					n = 2010				n = 2001				n = 1982			

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, SE = standard error, n=number of observations.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup> Total average income in thousands of dollars

### 5.4.2 Associations between Arsenic Concentration and Diabetes Prevalence

The Bayesian analysis did not identify any associations between arsenic concentrations in drinking water and diabetes prevalence for 2010 for public water supplies or private wells where the credible interval for the RR did not include 1 (Table 5.9). Similarly, there were no significant associations identified in any of the frequentist GLMMs for 2006-2010 (Tables 5.9 and 5.10).

**Table 5.9** Comparison of Bayesian and frequentist models of the association between arsenic concentrations in drinking water and diabetes prevalence for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		
		lower	upper			lower	upper	p
Intercept	1.024	0.957	1.096	Intercept	1.033	0.946	1.075	0.79
Arsenic (µg/L)	0.974	0.922	1.027	Arsenic (µg/L)	1.025	0.938	1.033	0.52
COPD prevalence	1.212	0.962	1.518	COPD prevalence	1.124	0.961	1.518	0.11
Education <sup>1</sup>	0.945	0.867	1.032	Education <sup>1</sup>	1.045	0.874	1.041	0.29
Income <sup>2</sup>	1.000	0.998	1.002	Income <sup>2</sup>	1.001	0.997	1.001	0.39
Random effects		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.064	0.020	Unstructured RE		0.013	0.002	
Spatially structured RE		0.093	0.016					
Burn in period = 160,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		
		lower	upper			lower	upper	p
Intercept	1.020	0.961	1.082	Intercept	1.031	0.942	1.061	0.99
Arsenic (µg/L)	0.985	0.955	1.016	Arsenic (µg/L)	1.015	0.966	1.024	0.72
COPD prevalence	1.210	0.962	1.517	COPD prevalence	1.124	0.960	1.517	0.11
Education <sup>1</sup>	0.946	0.867	1.032	Education <sup>1</sup>	1.045	0.875	1.042	0.30
Income <sup>2</sup>	1.000	0.998	1.002	Income <sup>2</sup>	1.001	0.997	1.001	0.39
Random effects variance		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.064	0.020	Unstructured RE		0.013	0.002	
Spatially structured RE		0.093	0.016					
Burn in period = 100,000 iterations								
Number of observations = 2114								

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

**Table 5.10** Summary of the results for frequentist GLMMs examining the association between arsenic concentrations in drinking water and diabetes prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006				
	95% CI				95% CI				95% CI				95% CI				
	Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.989	0.926	1.055	0.73	1.005	0.941	1.072	0.89	1.018	0.951	1.088	0.61	1.018	0.949	1.091	0.62	
Arsenic (µg/L)	0.989	0.941	1.038	0.64	0.979	0.932	1.028	0.40	0.971	0.924	1.021	0.25	0.972	0.922	1.024	0.28	
COPD prevalence	1.157	0.911	1.471	0.23	1.182	0.924	1.513	0.18	1.404	1.077	1.829	0.01	1.380	1.021	1.865	0.04	
Education <sup>1</sup>	1.002	0.916	1.095	0.97	0.987	0.901	1.081	0.78	0.948	0.862	1.042	0.26	0.953	0.863	1.051	0.34	
Income <sup>2</sup>	0.998	0.996	1.000	0.12	0.999	0.997	1.001	0.19	0.999	0.997	1.002	0.59	0.999	0.996	1.001	0.23	
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE		
	0.013		0.003		0.013		0.002		0.013		0.003		0.014		0.003		
<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI				
	Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
	Intercept	0.979	0.921	1.040	0.49	0.974	0.916	1.035	0.40	0.981	0.921	1.044	0.54	0.975	0.914	1.041	0.45
Arsenic (µg/L)	0.998	0.969	1.028	0.92	1.005	0.976	1.036	0.72	1.004	0.974	1.035	0.82	1.008	0.977	1.040	0.63	
COPD prevalence	1.157	0.910	1.471	0.23	1.182	0.923	1.512	0.19	1.405	1.078	1.830	0.01	1.381	1.022	1.867	0.04	
Education <sup>1</sup>	1.002	0.916	1.096	0.96	0.988	0.902	1.082	0.79	0.949	0.863	1.043	0.28	0.954	0.864	1.053	0.35	
Income <sup>2</sup>	0.998	0.996	1.000	0.12	0.999	0.997	1.001	0.18	0.999	0.997	1.001	0.56	0.999	0.996	1.001	0.22	
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE		
	0.013		0.003		0.013		0.003		0.013		0.003		0.014		0.003		
n=2090					n = 2085				n = 2064				n = 2046				

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, SE = standard error, n=number of observations.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Total average income in thousands of dollars

### 5.4.3 Associations between Health Standards and Diabetes Incidence

There were no associations detected between and health standards PC scores for either public water supplies or private wells and diabetes incidence from 2010-2012 (Table 5.11).

**Table 5.11** Comparison of Bayesian and frequentist models of the association between drinking water health standards principal component scores and cumulative diabetes incidence 2010-2012 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>		95% CrI				95% CI		
Effect	RR	lower	upper	Effect	RR	lower	upper	p
Intercept	0.988	0.925	1.055	Intercept	0.989	0.926	1.056	0.74
PC1 <sub>health</sub>	1.069	0.941	1.221	PC1 <sub>health</sub>	1.047	0.941	1.165	0.40
PC2 <sub>health</sub>	0.967	0.868	1.075	PC2 <sub>health</sub>	0.965	0.879	1.060	0.46
PC3 <sub>health</sub>	0.952	0.860	1.054	PC3 <sub>health</sub>	0.965	0.881	1.058	0.45
COPD prevalence	1.812	1.051	3.080	COPD prevalence	1.860	1.092	3.169	0.02
Education <sup>1</sup>	0.886	0.742	1.058	Education <sup>1</sup>	0.884	0.741	1.054	0.17
Income <sup>2</sup>	0.998	0.995	1.001	Income <sup>2</sup>	0.998	0.995	1.001	0.24
Random effects		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.081	0.031	Unstructured RE		0.012	0.005	
Spatially structured RE		0.075	0.032					
Burn in period = 100,000 iterations								
<i>Private wells</i>		95% CrI				95% CI		
Effect	RR	lower	upper	Effect	RR	lower	upper	p
Intercept	0.989	0.924	1.058	Intercept	0.989	0.924	1.059	0.76
PC1 <sub>health</sub>	1.010	0.905	1.127	PC1 <sub>health</sub>	1.019	0.919	1.130	0.72
PC2 <sub>health</sub>	1.040	0.955	1.132	PC2 <sub>health</sub>	1.044	0.966	1.128	0.28
PC3 <sub>health</sub>	1.020	0.889	1.172	PC3 <sub>health</sub>	1.018	0.891	1.162	0.80
COPD prevalence	1.845	1.070	3.121	COPD prevalence	1.856	1.088	3.165	0.02
Education <sup>1</sup>	0.882	0.739	1.055	Education <sup>1</sup>	0.881	0.739	1.051	0.16
Income <sup>2</sup>	0.998	0.994	1.001	Income <sup>2</sup>	0.998	0.994	1.001	0.21
Random effects		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.096	0.029	Unstructured RE		0.013	0.005	
Spatially structured RE		0.060	0.031					
Burn in period = 160,000 iterations								
Number of observations = 2041								

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).



In the frequentist GLMMs for the previous four time periods for public water supplies, increased PC1<sub>health</sub> score was associated with an increase in diabetes incidence only for 2007-2009 (RR=1.108, 95% CI 1.002-1.226); no other associations were evident (Table 5.12). For private water supplies, an increase in the PC2<sub>health</sub> score was associated with increased diabetes incidence only for 2008-2010 (RR=1.095, 95% CI 1.021-1.174); no other associations were demonstrated.

**Table 5.12** Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and 3-year cumulative incidence of diabetes for the time periods 2006-2008 through 2009-2011 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009 - 2011				2008 – 2010				2007 - 2009				2006 - 2008			
	Effect	95% CI			RR	95% CI			RR	95% CI			RR	95% CI		
		RR	lower	upper		p	lower	upper		p	lower	upper		p	lower	upper
Intercept	0.988	0.930	1.051	0.71	0.983	0.925	1.044	0.57	0.975	0.917	1.037	0.43	0.969	0.909	1.032	0.33
PC1 <sub>health</sub>	1.049	0.952	1.155	0.33	1.082	0.983	1.190	0.11	<b>1.108</b>	<b>1.002</b>	<b>1.226</b>	<b>0.045</b>	1.112	0.995	1.242	0.06
PC2 <sub>health</sub>	0.977	0.898	1.064	0.60	0.932	0.856	1.014	0.10	0.993	0.909	1.084	0.88	0.951	0.865	1.047	0.31
PC3 <sub>health</sub>	1.013	0.932	1.100	0.77	1.034	0.952	1.124	0.43	1.023	0.938	1.115	0.61	0.982	0.894	1.079	0.71
COPD prevalence	1.732	1.043	2.876	0.03	1.398	0.823	2.372	0.22	1.415	0.811	2.470	0.22	1.200	0.656	2.196	0.55
Education <sup>1</sup>	0.917	0.777	1.083	0.31	0.971	0.824	1.143	0.72	0.976	0.829	1.150	0.77	1.017	0.865	1.197	0.84
Income <sup>2</sup>	0.998	0.995	1.001	0.24	0.999	0.996	1.002	0.45	0.997	0.994	1.001	0.11	0.999	0.996	1.002	0.52
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.007		0.004		0.007		0.004		0.011		0.005		0.020		0.005	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	Effect	95% CI			RR	95% CI			RR	95% CI			RR	95% CI		
		RR	lower	upper		p	lower	upper		p	lower	upper		p	lower	upper
Intercept	0.997	0.936	1.062	0.93	1.004	0.943	1.069	0.89	1.002	0.940	1.068	0.95	0.981	0.919	1.048	0.57
PC1 <sub>health</sub>	1.034	0.942	1.135	0.49	1.029	0.936	1.131	0.55	0.943	0.856	1.040	0.24	0.968	0.871	1.076	0.55
PC2 <sub>health</sub>	1.060	0.989	1.137	0.10	<b>1.095</b>	<b>1.021</b>	<b>1.174</b>	<b>0.01</b>	1.060	0.987	1.139	0.11	1.026	0.948	1.111	0.53
PC3 <sub>health</sub>	0.996	0.883	1.122	0.94	0.923	0.818	1.041	0.19	0.901	0.795	1.020	0.10	0.947	0.825	1.087	0.44
COPD prevalence	1.750	1.054	2.908	0.03	1.433	0.845	2.432	0.18	1.452	0.834	2.528	0.19	1.202	0.657	2.199	0.55
Education <sup>1</sup>	0.907	0.768	1.070	0.25	0.956	0.812	1.126	0.59	0.957	0.813	1.127	0.60	1.015	0.862	1.194	0.86
Income <sup>2</sup>	0.998	0.995	1.001	0.20	0.998	0.995	1.002	0.31	0.997	0.994	1.000	0.07	0.999	0.995	1.002	0.41
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.008		0.004		0.008		0.004		0.011		0.005		0.020		0.005	
	n=2024				n = 2010				n = 2001				n = 1982			

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error, n=number of observations.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

#### *5.4.4 Associations between Health Standards and Diabetes Prevalence*

From the Bayesian model for public water supplies that included all PC<sub>health</sub> scores as continuous variables, as the score for the first PC<sub>health</sub> increased, the prevalence of diabetes increased (RR=1.125, 95% CrI 1.027-1.236). The probability that the RR > 1 for the coefficient for PC1<sub>health</sub> was 99.4% for public water supplies.

When the first PC<sub>health</sub> score for public water supplies was subsequently categorized into quintiles, the relationship between PC1<sub>health</sub> and diabetes prevalence was not linear in the resulting Bayesian model (Table 5.13). Compared to the lowest quintile, having a PC1<sub>health</sub> score in the third quintile (RR=1.101, 95% CrI 1.019-1.188), fourth quintile (RR=1.088, 95% CrI 1.003-1.180) or fifth quintile (RR=1.115, 95% CrI 1.026-1.213) for public water supplies was associated with an increase in diabetes prevalence (Table 5.13). The probability that the rate ratio for each quintile compared to the first quintile > 1 was 91.6% for the second quintile, 99.3% for the third quintile, 97.9% for the fourth quintile, and 99.5% for the highest quintile.

In the frequentist GLMM for 2010 with health standards PC scores as continuous variables for public water supplies, PC1<sub>health</sub> was not significantly associated with diabetes prevalence (RR=1.072, 95% CI 0.996-1.154, p=0.06). With the PC1<sub>health</sub> score categorized into quintiles for direct comparison to the Bayesian model, having a score in the third quintile compared to the first was associated with an increase in diabetes prevalence (RR=1.089, 95% CI 1.011-1.173). No differences in diabetes prevalence were evident when comparing the remaining quintiles of PC1<sub>health</sub> to the lowest quintile (Table 5.13), and the type III test of fixed effects (p=0.22) for the categorized form of PC1<sub>health</sub> was not significant.

Also in the frequentist model, increasing scores for PC2<sub>health</sub> in public water supplies were associated with a decrease in prevalence of diabetes in 2010 (RR=0.931, 95% CI 0.873- 0.992); however, in the Bayesian model the credible interval for PC2<sub>health</sub> included 1 (RR=0.930, 95% CrI 0.859-1.007).

There were no associations identified between health standards PC scores in private wells and diabetes prevalence for 2010 (Table 5.13).

**Table 5.13** Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and diabetes prevalence for 2010 in rural southern Saskatchewan.

Bayesian model					Frequentist model				
<i>Public water supplies</i>									
Effect	RR	95% CrI			Effect	RR	95% CI		p
		lower	upper				lower	upper	
Intercept	0.932	0.875	0.993		Intercept	0.946	0.892	1.004	0.07
PC1 <sub>health</sub> Quintile 1	reference category				PC1 <sub>health</sub> Quintile 1	reference category			
PC1 <sub>health</sub> Quintile 2	1.052	0.979	1.131		PC1 <sub>health</sub> Quintile 2	1.045	0.971	1.124	0.24
PC1 <sub>health</sub> Quintile 3	<b>1.101</b>	<b>1.019</b>	<b>1.188</b>		PC1 <sub>health</sub> Quintile 3	<b>1.089</b>	<b>1.011</b>	<b>1.173</b>	<b>0.03</b>
PC1 <sub>health</sub> Quintile 4	<b>1.088</b>	<b>1.003</b>	<b>1.180</b>		PC1 <sub>health</sub> Quintile 4	1.062	0.984	1.146	0.12
PC1 <sub>health</sub> Quintile 5	<b>1.115</b>	<b>1.026</b>	<b>1.213</b>		PC1 <sub>health</sub> Quintile 5	1.070	0.994	1.151	0.07
PC2 <sub>health</sub>	0.930	0.859	1.007		PC2 <sub>health</sub>	<b>0.931</b>	<b>0.873</b>	<b>0.992</b>	<b>0.03</b>
PC3 <sub>health</sub>	0.970	0.903	1.042		PC3 <sub>health</sub>	0.980	0.922	1.042	0.53
COPD prevalence	1.192	0.947	1.500		COPD prevalence	1.213	0.966	1.524	0.10
Education <sup>1</sup>	0.950	0.871	1.037		Education <sup>1</sup>	0.951	0.872	1.038	0.26
Income <sup>2</sup>	1.000	0.998	1.002		Income <sup>2</sup>	0.999	0.997	1.001	0.49
Random effects variance		SD	SE		Random Effect		Variance	SE	
Unstructured RE		0.060	0.019		Unstructured RE		0.012	0.002	
Spatially structured RE		0.095	0.016						
Burn in period = 60,000 iterations									
<i>Private wells</i>									
Effect	RR	95% CrI			Effect	RR	95% CI		p
		lower	upper				lower	upper	
Intercept	1.007	0.970	1.046		Intercept	1.006	0.966	1.047	0.77
PC1 <sub>health</sub>	0.963	0.892	1.042		PC1 <sub>health</sub>	0.942	0.880	1.009	0.09
PC2 <sub>health</sub>	1.034	0.973	1.098		PC2 <sub>health</sub>	1.035	0.983	1.091	0.19
PC3 <sub>health</sub>	0.936	0.852	1.028		PC3 <sub>health</sub>	0.928	0.849	1.014	0.10
COPD prevalence	1.222	0.974	1.534		COPD prevalence	1.212	0.965	1.523	0.10
Education <sup>1</sup>	0.945	0.865	1.030		Education <sup>1</sup>	0.951	0.872	1.037	0.26
Income <sup>2</sup>	0.999	0.997	1.001		Income <sup>2</sup>	0.999	0.997	1.001	0.33
Random effects		SD	SE		Random effects		Variance	SE	
Unstructured RE		0.071	0.020		Unstructured RE		0.012	0.002	
Spatially structured RE		0.083	0.019						
Burn in period = 120,000 iterations									
Number of observations = 2114									

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

<sup>3</sup>Overall p value for categorized variable based on type III likelihood ratio test.

Repeating the analysis using frequentist GLMMs for each year from 2006 through 2009 did not identify any significant associations between the first or third PC<sub>health</sub> scores in public water supplies and the prevalence of diabetes (Table 5.14). However, the association between increasing PC<sub>2health</sub> score in public water supplies and decreased prevalence was significant in 2006 (RR=0.932, 95% CI 0.870-0.997), 2007 (RR=0.925, 95% CI 0.867-0.988), and 2009 (RR=0.924, 95% CI 0.867-0.985), but not 2008 (Table 5.14).

The only significant association between PC<sub>health</sub> scores in the private well data and diabetes prevalence was for PC<sub>1health</sub> in 2008 (Table 5.14), where an increase in the PC<sub>1health</sub> scores was associated with a decrease in diabetes prevalence (RR=0.928, 95% CI 0.865-0.995).

**Table 5.14** Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and diabetes prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006			
	95% CI				95% CI				95% CI				95% CI			
	Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper
Intercept	0.978	0.940	1.017	0.27	0.983	0.945	1.024	0.41	0.990	0.950	1.031	0.62	0.990	0.949	1.033	0.64
PC1 <sub>health</sub>	1.063	0.986	1.145	0.11	1.053	0.977	1.136	0.18	1.036	0.959	1.119	0.37	1.029	0.950	1.115	0.49
PC2 <sub>health</sub>	<b>0.924</b>	<b>0.867</b>	<b>0.985</b>	<b>0.02</b>	0.944	0.885	1.006	0.08	<b>0.925</b>	<b>0.867</b>	<b>0.988</b>	<b>0.02</b>	<b>0.932</b>	<b>0.870</b>	<b>0.997</b>	<b>0.04</b>
PC3 <sub>health</sub>	0.976	0.916	1.039	0.44	0.972	0.912	1.036	0.38	0.969	0.908	1.034	0.34	0.978	0.915	1.046	0.52
COPD prevalence	1.158	0.911	1.471	0.23	1.181	0.923	1.512	0.19	1.402	1.076	1.826	0.01	1.380	1.021	1.864	0.04
Education <sup>1</sup>	0.998	0.912	1.091	0.96	0.984	0.898	1.078	0.73	0.944	0.859	1.037	0.23	0.950	0.860	1.048	0.30
Income <sup>2</sup>	0.999	0.997	1.001	0.16	0.999	0.997	1.001	0.23	1.000	0.997	1.002	0.67	0.999	0.997	1.001	0.27
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.012		0.002		0.012		0.002		0.012		0.003		0.013		0.003	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI				
	Effect	RR	95% CI		p	RR	95% CI		p	RR	95% CI		p	RR	95% CI		p
			lower	upper			lower	upper			lower	upper			lower	upper	
Intercept	0.985	0.946	1.027	0.49	0.986	0.946	1.028	0.52	0.994	0.952	1.038	0.78	0.995	0.952	1.041	0.84	
PC1 <sub>health</sub>	0.941	0.877	1.009	0.09	<b>0.928</b>	<b>0.865</b>	<b>0.995</b>	<b>0.04</b>	0.950	0.884	1.022	0.17	0.944	0.875	1.018	0.13	
PC2 <sub>health</sub>	1.034	0.981	1.091	0.22	1.009	0.956	1.064	0.75	1.024	0.968	1.082	0.41	1.022	0.965	1.082	0.46	
PC3 <sub>health</sub>	0.964	0.880	1.056	0.43	0.983	0.897	1.077	0.71	0.965	0.878	1.061	0.46	0.960	0.871	1.058	0.41	
COPD prevalence	1.159	0.912	1.472	0.23	1.181	0.923	1.511	0.19	1.404	1.078	1.829	0.01	1.377	1.019	1.861	0.04	
Education <sup>1</sup>	0.998	0.913	1.091	0.97	0.985	0.899	1.079	0.75	0.946	0.861	1.040	0.25	0.951	0.862	1.050	0.32	
Income <sup>2</sup>	0.998	0.996	1.000	0.09	0.998	0.996	1.001	0.15	0.999	0.997	1.001	0.51	0.999	0.996	1.001	0.19	
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE		
	0.013		0.002		0.012		0.002		0.013		0.003		0.014		0.003		
	n=2090				n = 2085				n = 2064				n = 2046				

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error, n=number of observations.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

#### *5.4.5 Associations between Aesthetic Objective Factor Scores and Diabetes Incidence*

No associations were found between aesthetic objectives PC scores for public water supplies and diabetes incidence for 2010-2012 in either the Bayesian or frequentist analysis (Table 5.15).

In the Bayesian model examining associations between aesthetic objectives PC scores from private wells and diabetes incidence for 2010-2012, an increase in the  $PC3_{\text{aesthetic}}$  score was associated with a decrease in diabetes incidence ( $RR=0.863$ , 95% CrI 0.746-0.998). However for private well data, there were no differences in diabetes incidence between the quintiles two through five and quintile one of  $PC3_{\text{aesthetic}}$  where 1 was not included in the CrI (Table 5.15). No significant associations were identified between  $PC_{\text{aesthetic}}$  scores and diabetes incidence in the frequentist GLMM for 2010-2012 ( $RR=0.884$ , 95% CI 0.771-1.013,  $p=0.08$ ).  $PC3_{\text{aesthetic}}$  for private wells was categorized for comparison to the Bayesian model. The type III test of fixed effects for the categorized  $PC3_{\text{aesthetic}}$  was not significant ( $p=0.51$ ) and none of the differences in diabetes incidence between increasing quintiles and the lowest category were significant (Table 5.15).

**Table 5.15** Comparison of Bayesian and frequentist models of the association between drinking water aesthetic objectives principal component scores and diabetes cumulative incidence 2010-2012 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.987	0.925	1.054	Intercept	0.989	0.927	1.056	0.74
PC1 <sub>aesthetic</sub>	1.084	0.977	1.201	PC1 <sub>aesthetic</sub>	1.087	0.989	1.194	0.08
PC2 <sub>aesthetic</sub>	1.049	0.960	1.145	PC2 <sub>aesthetic</sub>	1.053	0.978	1.134	0.17
PC3 <sub>aesthetic</sub>	1.018	0.907	1.136	PC3 <sub>aesthetic</sub>	1.043	0.948	1.148	0.38
PC4 <sub>aesthetic</sub>	1.039	0.917	1.180	PC4 <sub>aesthetic</sub>	1.039	0.923	1.170	0.53
COPD prevalence	1.789	1.044	3.037	COPD prevalence	1.817	1.066	3.097	0.03
Education <sup>1</sup>	0.891	0.746	1.065	Education <sup>1</sup>	0.890	0.746	1.062	0.20
Income <sup>2</sup>	0.998	0.995	1.002	Income <sup>2</sup>	0.998	0.995	1.002	0.31
Random effects		SD		Random Effect		Variance		SE
Unstructured RE		0.087		Unstructured RE		0.011		0.005
Spatially structured RE		0.061						
Burn in period = 120,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.048	0.946	1.060	Intercept	1.047	0.949	1.155	0.36
PC1 <sub>aesthetic</sub>	0.924	0.844	1.012	PC1 <sub>aesthetic</sub>	0.926	0.852	1.005	0.07
PC2 <sub>aesthetic</sub>	0.955	0.878	1.040	PC2 <sub>aesthetic</sub>	0.958	0.885	1.036	0.28
PC3 <sub>aesthetic</sub> Quintile 1	Reference category			PC3 <sub>aesthetic</sub> Quintile 1	Reference category			0.51 <sup>3</sup>
PC3 <sub>aesthetic</sub> Quintile 2	0.974	0.865	1.097	PC3 <sub>aesthetic</sub> Quintile 2	0.967	0.864	1.084	0.57
PC3 <sub>aesthetic</sub> Quintile 3	0.956	0.852	1.073	PC3 <sub>aesthetic</sub> Quintile 3	0.955	0.855	1.067	0.41
PC3 <sub>aesthetic</sub> Quintile 4	0.913	0.812	1.026	PC3 <sub>aesthetic</sub> Quintile 4	0.928	0.812	1.015	0.09
PC3 <sub>aesthetic</sub> Quintile 5	0.912	0.806	1.033	PC3 <sub>aesthetic</sub> Quintile 5	0.929	0.828	1.043	0.21
COPD prevalence	1.871	1.081	3.187	COPD prevalence	1.896	1.112	3.232	0.02
Education <sup>1</sup>	0.882	0.739	1.052	Education <sup>1</sup>	0.880	0.738	1.049	0.16
Income <sup>2</sup>	0.997	0.994	1.001	Income <sup>2</sup>	0.997	0.994	1.001	0.13
Random effects		SD		Random Effect		Variance		SE
Unstructured RE		0.095		Unstructured RE		0.011		0.005
Spatially structured RE		0.062						
Burn in period = 120,000 iterations								
Number of observations = 2041								

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

<sup>3</sup>Overall p value for categorized variable based on type III likelihood ratio test.



From the frequentist analysis of public water supply data, there were no associations between any of the aesthetic objectives PC scores and diabetes incidence summarized for the three-year time periods starting from 2006 to 2009 (Table 5.16). For the private well data, an increase in the score for PC1<sub>aesthetic</sub> was associated with decreased diabetes incidence in 2008-2010 (RR=0.906, 95% CI 0.839-0.978) and 2009–2011 (RR=0.909, 95% CI 0.843-0.980), but not in earlier time periods. No other associations were identified between aesthetic objectives PC scores in private wells and diabetes incidence (Table 5.16).

**Table 5.16** Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and 3-year cumulative incidence of diabetes for the time periods 2006-2008 through 2009-2011 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009-2011				2008-2010				2007-2009				2006-2008			
	Effect	95% CI			RR	95% CI			RR	95% CI			RR	95% CI		
		RR	lower	upper		p	lower	upper		p	lower	upper		p	lower	upper
Intercept	0.990	0.932	1.052	0.76	0.984	0.926	1.045	0.60	0.978	0.920	1.039	0.47	0.970	0.911	1.032	0.34
PC1 <sub>aesthetic</sub>	1.033	0.949	1.124	0.46	1.006	0.923	1.097	0.90	1.010	0.924	1.104	0.83	1.025	0.929	1.130	0.63
PC2 <sub>aesthetic</sub>	1.064	0.995	1.138	0.07	1.070	0.999	1.145	0.054	1.037	0.966	1.113	0.32	1.019	0.943	1.100	0.64
PC3 <sub>aesthetic</sub>	1.025	0.941	1.116	0.57	1.009	0.925	1.101	0.83	0.959	0.875	1.050	0.37	1.005	0.909	1.111	0.92
PC4 <sub>aesthetic</sub>	1.053	0.947	1.171	0.34	1.084	0.973	1.208	0.14	1.107	0.989	1.239	0.08	1.127	0.996	1.275	0.06
COPD prevalence	1.665	1.002	2.767	0.049	1.323	0.777	2.253	0.30	1.348	0.771	2.359	0.30	1.156	0.631	2.118	0.64
Education <sup>1</sup>	0.923	0.782	1.090	0.34	0.982	0.833	1.157	0.83	1.988	0.839	1.164	0.89	1.030	0.875	1.212	0.73
Income <sup>2</sup>	0.998	0.995	1.001	0.30	0.999	0.996	1.002	0.47	0.997	0.994	1.001	0.10	0.999	0.995	1.002	0.41
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.007		0.004		0.007		0.004		0.011		0.005		0.019		0.005	
<i>Private wells</i>	Effect	95% CI			RR	95% CI			RR	95% CI			RR	95% CI		
		RR	lower	upper		p	lower	upper		p	lower	upper		p	lower	upper
	Intercept	1.005	0.945	1.069	0.87	0.999	0.940	1.063	0.98	0.991	0.931	1.055	0.78	0.972	0.912	1.037
PC1 <sub>aesthetic</sub>	<b>0.909</b>	<b>0.843</b>	<b>0.980</b>	<b>0.01</b>	<b>0.906</b>	<b>0.839</b>	<b>0.978</b>	<b>0.01</b>	0.935	0.863	1.013	0.10	0.993	0.910	1.084	0.88
PC2 <sub>aesthetic</sub>	0.989	0.923	1.060	0.76	0.996	0.928	1.069	0.90	0.985	0.915	1.060	0.68	0.989	0.914	1.070	0.78
PC3 <sub>aesthetic</sub>	0.889	0.786	1.007	0.06	0.902	0.795	1.024	0.11	0.955	0.836	1.091	0.50	1.008	0.872	1.165	0.91
COPD prevalence	1.758	1.060	2.916	0.03	1.416	0.834	2.404	0.20	1.431	0.820	2.496	0.21	1.194	0.652	2.187	0.57
Education <sup>1</sup>	0.906	0.758	1.069	0.24	0.959	0.814	1.129	0.61	0.961	0.816	1.132	0.64	1.018	0.865	1.198	0.83
Income <sup>2</sup>	0.998	0.995	1.001	0.15	0.998	0.995	1.001	0.29	0.997	0.994	1.000	0.08	0.999	0.995	1.002	0.42
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.007		0.004		0.008		0.004		0.012		0.005		0.020		0.006	
n=2024					n = 2010				n = 2001				n = 1982			

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error, n=number of observations.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

#### *5.4.6 Associations between Aesthetic Objective Factor Scores and Diabetes Prevalence*

No associations were identified between aesthetic objectives PC scores in either public water supplies or private wells and diabetes prevalence in 2010 from either the Bayesian or frequentist models (Table 5.17). Similarly, frequentist GLMMs for each year from 2006–2009 did not suggest any significant associations between aesthetic objective PC scores and diabetes prevalence for either type of water supply (Table 5.18).

**Table 5.17** Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and diabetes prevalence for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.993	0.959	1.029	Intercept	0.991	0.954	1.029	0.63
PC1 <sub>aesthetic</sub>	1.019	0.943	1.102	PC1 <sub>aesthetic</sub>	1.005	0.942	1.073	0.88
PC2 <sub>aesthetic</sub>	1.015	0.946	1.090	PC2 <sub>aesthetic</sub>	1.034	0.982	1.089	0.20
PC3 <sub>aesthetic</sub>	0.960	0.882	1.042	PC3 <sub>aesthetic</sub>	1.008	0.942	1.079	0.81
PC4 <sub>aesthetic</sub>	1.050	0.958	1.152	PC4 <sub>aesthetic</sub>	1.048	0.965	1.138	0.27
COPD prevalence	1.199	0.953	1.504	COPD prevalence	1.192	0.948	1.499	0.13
Education <sup>1</sup>	0.949	0.870	1.036	Education <sup>1</sup>	0.957	0.877	1.045	0.33
Income <sup>2</sup>	1.000	0.998	1.002	Income <sup>2</sup>	0.999	0.997	1.001	0.45
Random effects variance		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.059	0.020	Unstructured RE		0.013	0.002	
Spatially structured RE		0.101	0.017					
Burn in period = 80,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.005	0.968	1.042	Intercept	0.996	0.957	1.035	0.83
PC1 <sub>aesthetic</sub>	0.964	0.903	1.029	PC1 <sub>aesthetic</sub>	0.967	0.912	1.025	0.26
PC2 <sub>aesthetic</sub>	0.964	0.907	1.024	PC2 <sub>aesthetic</sub>	0.975	0.926	1.027	0.34
PC3 <sub>aesthetic</sub>	0.939	0.847	1.041	PC3 <sub>aesthetic</sub>	1.007	0.916	1.107	0.88
COPD prevalence	1.215	0.965	1.527	COPD prevalence	1.205	0.959	1.515	0.11
Education <sup>1</sup>	0.945	0.866	1.031	Education <sup>1</sup>	0.953	0.873	1.040	0.28
Income <sup>2</sup>	1.000	0.998	1.001	Income <sup>2</sup>	0.999	0.997	1.001	0.35
Random effects variance		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.059	0.019	Unstructured RE		0.013	0.002	
Spatially structured RE		0.099	0.016					
Burn in period = 80,000 iterations								
Number of observations = 2114								

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, PC = principal component, SD = standard deviation SE = standard error, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

**Table 5.18** Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and diabetes prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006			
	95% CI				95% CI				95% CI				95% CI			
	Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper
Intercept	0.976	0.939	1.015	0.23	0.982	0.944	1.022	0.38	0.986	0.947	1.027	0.51	0.987	0.946	1.030	0.55
PC1 <sub>aesthetic</sub>	0.999	0.935	1.068	0.98	0.993	0.929	1.062	0.85	0.993	0.927	1.064	0.85	0.989	0.921	1.062	0.76
PC2 <sub>aesthetic</sub>	1.032	0.979	1.088	0.24	1.019	0.967	1.074	0.48	1.033	0.979	1.090	0.24	1.035	0.979	1.095	0.23
PC3 <sub>aesthetic</sub>	1.028	0.959	1.102	0.43	1.020	0.952	1.094	0.57	1.027	0.956	1.102	0.47	1.018	0.946	1.096	0.63
PC4 <sub>aesthetic</sub>	1.044	0.960	1.136	0.32	1.038	0.954	1.129	0.39	1.017	0.932	1.109	0.71	1.007	0.920	1.102	0.88
COPD prevalence	1.146	0.901	1.457	0.27	1.171	0.914	1.500	0.21	1.392	1.068	1.815	0.01	1.368	1.012	1.850	0.04
Education <sup>1</sup>	1.003	0.917	1.097	0.95	0.989	0.902	1.083	0.81	0.948	0.863	1.042	0.27	0.953	0.863	1.052	0.34
Income <sup>2</sup>	0.999	0.996	1.001	0.15	0.999	0.997	1.001	0.20	1.000	0.997	1.002	0.67	0.999	0.997	1.001	0.29
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.013		0.003		0.012		0.002		0.013		0.003		0.014		0.003	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	95% CI				95% CI				95% CI				95% CI			
	Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper
Intercept	0.981	0.942	1.021	0.34	0.983	0.944	1.024	0.41	0.987	0.947	1.029	0.55	0.989	0.947	1.033	0.62
PC1 <sub>aesthetic</sub>	0.970	0.914	1.029	0.31	0.989	0.932	1.050	0.71	0.977	0.918	1.038	0.45	0.971	0.912	1.035	0.37
PC2 <sub>aesthetic</sub>	0.971	0.920	1.023	0.27	0.969	0.919	1.022	0.25	0.982	0.930	1.038	0.52	0.979	0.925	1.036	0.47
PC3 <sub>aesthetic</sub>	1.008	0.915	1.110	0.88	1.040	0.944	1.146	0.42	1.033	0.935	1.141	0.52	1.029	0.929	1.141	0.58
COPD prevalence	1.157	0.910	1.471	0.23	1.182	0.923	1.513	0.19	1.402	1.076	1.826	0.01	1.376	1.018	1.860	0.04
Education <sup>1</sup>	1.000	0.914	1.093	0.99	0.986	0.900	1.080	0.76	0.947	0.861	1.040	0.25	0.951	0.862	1.050	0.32
Income <sup>2</sup>	0.998	0.996	1.000	0.10	0.999	0.997	1.001	0.17	0.999	0.997	1.001	0.56	0.999	0.996	1.001	0.21
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.013		0.003		0.012		0.002		0.013		0.003		0.014		0.003	
	n=2090				n = 2085				n = 2064				n = 2046			

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error, n=number of observations.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

### 5.4.7 Summary of results

The effect estimates for the Bayesian models evaluating associations between all of the water quality risk factors (as continuous variables only) and diabetes incidence in 2010-2012 and diabetes prevalence in 2010 are summarized in Table 5.19.

**Table 5.19** Summary of effect estimates from Bayesian analysis of associations between drinking water risk factors from public water supply and private well data and cumulative incidence of diabetes for 2010-2012 and prevalence for 2010. Results in bold type indicate that the 95% credible interval for that effect estimate did not include 1.

Risk Factor	Incidence		Prevalence	
	Public Water	Private Wells	Public Water	Private Wells
	RR (95% CrI)	RR (95% CrI)	RR (95% CrI)	RR (95% CrI)
<b>Arsenic</b>	0.980 (0.910-1.054)	<b>0.954 (0.913-0.997)</b>	0.974 (0.922-1.027)	0.985 (0.955-1.016)
<i>Health Standards</i>				
<b>PC1<sub>health</sub></b>	1.069 (0.941-1.221)	1.010 (0.905-1.127)	<b>1.125 (1.027-1.236)</b>	0.963 (0.892-1.042)
<b>PC2<sub>health</sub></b>	0.967 (0.868-1.075)	1.040 (0.955-1.132)	0.930 (0.859-1.007)	1.034 (0.973-1.098)
<b>PC3<sub>health</sub></b>	0.952 (0.860-1.054)	1.020 (0.889-1.172)	0.970 (0.903-1.042)	0.936 (0.852-1.028)
<i>Aesthetic Objectives</i>				
<b>PC1<sub>aesthetic</sub></b>	1.084 (0.977-1.201)	0.918 (0.838-1.006)	1.019 (0.943-1.102)	0.964 (0.903-1.029)
<b>PC2<sub>aesthetic</sub></b>	1.049 (0.960-1.145)	0.952 (0.879-1.033)	1.015 (0.946-1.090)	0.964 (0.907-1.024)
<b>PC3<sub>aesthetic</sub></b>	1.018 (0.907-1.136)	<b>0.863 (0.746-0.998)</b>	0.960 (0.882-1.042)	0.939 (0.847-1.041)
<b>PC4<sub>aesthetic</sub></b>	1.039 (0.917-1.180)	–	1.050 (0.958-1.152)	–

Estimates are based on the continuous form of each risk factor variable and are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts, and by prevalence of COPD, education level and income by inclusion as risk factors in model.

RR = risk ratio, CrI = credible interval, PC = principal component.

The number of positive and negative associations along with the number of non-significant results for the frequentist GLMMs across five three-year time periods for diabetes incidence and 5 years for diabetes prevalence are summarized in Table 5.20.

**Table 5.20** Total number of non-significant results and positive and negative associations between the water quality risk factors and diabetes incidence or prevalence from the frequentist GLMM analysis over 5 time periods.

	Incidence						Prevalence					
	Public Water			Private wells			Public Water			Private wells		
	NS	Pos	Neg	NS	Pos	Neg	NS	Pos	Neg	NS	Pos	Neg
<b>Arsenic</b>	5	0	0	4	0	<b>1</b>	5	0	0	5	0	0
<i>Health Standards</i>												
<b>PC1<sub>health</sub></b>	4	<b>1</b>	0	5	0	0	5	0	0	4	0	<b>1</b>
<b>PC2<sub>health</sub></b>	5	0	0	4	<b>1</b>	0	1	0	<b>4</b>	5	0	0
<b>PC3<sub>health</sub></b>	5	0	0	5	0	0	5	0	0	5	0	0
<i>Aesthetic Objectives</i>												
<b>PC1<sub>aesthetic</sub></b>	5	0	0	3	0	<b>2</b>	5	0	0	5	0	0
<b>PC2<sub>aesthetic</sub></b>	5	0	0	5	0	0	5	0	0	5	0	0
<b>PC3<sub>aesthetic</sub></b>	5	0	0	5	0	0	5	0	0	5	0	0
<b>PC4<sub>aesthetic</sub></b>	5	0	0	–	–	–	5	0	0	–	–	–
<b>Total significant associations</b>	<b>1</b>		<b>0</b>	<b>1</b>		<b>3</b>	<b>0</b>		<b>4</b>	<b>0</b>		<b>1</b>

NS = not significant, Pos=positive association, Neg=negative association, PC=principal component

Given that 15 risk factors (8 for public water supplies and 7 for private wells) were assessed for associations with diabetes incidence or prevalence in the Bayesian models, the probability of falsely identifying an association between one water-related risk factor and prevalence (Table 5.21) was 53.7% at a level of significance of 0.05, while the probability of identifying the two associations between water-related exposures and cumulative incidences (Table 5.21) based on chance alone was calculated as 17.1%.

#### 5.4.8 Residuals and Model Fit

Analysis of the residuals from the frequentist models indicated that there was significant, although mild, spatial autocorrelation in the unexplained variance by region for the models examining associations between the arsenic concentration (Moran's  $I=0.029$ ,  $p=0.04$ ) and health standards PCs (Moran's  $I=0.072$ ,  $p=0.045$ ) and diabetes incidence. However, Moran's  $I$  was significant for all of the models examining associations between water quality variables and diabetes prevalence (Table 5.21).

**Table 5.21** Values for Moran's  $I$  for global spatial autocorrelation for residuals from frequentist models for diabetes incidence and prevalence from public and private water supplies. Risk factors include arsenic concentrations, health standards PC scores, or aesthetic objective PC scores as risk factors. Each model was adjusted for prevalence of COPD, education level and income. Residuals for null models or models for diabetes incidence and prevalence with no fixed effects were also analyzed.

Model	Moran's $I$	p
Diabetes Incidence		
<i>Public water supplies</i>		
Arsenic	0.069	0.04
Health Standards PCs	0.072	0.045
Aesthetic Objective PCs	0.046	0.12
<i>Private wells</i>		
Arsenic	0.057	0.12
Health Standards PCs	0.053	0.13
Aesthetic Objective PCs	0.040	0.17
Null model	0.077	0.04
Diabetes Prevalence		
<i>Public water supplies</i>		
Arsenic	0.176	0.001
Health Standards PCs	0.158	0.001
Aesthetic Objective PCs	0.159	0.002
<i>Private wells</i>		
Arsenic	0.175	0.001
Health Standards PCs	0.144	0.003
Aesthetic Objective PCs	0.170	0.002
Null model	0.178	0.001



For the model of associations between arsenic concentrations in private well water and diabetes incidence with both spatially structured and unstructured random effects the DIC was 5875, compared to a DIC of 5890 for the model with no spatially structured random effect.

For the model of the associations between diabetes prevalence and health standards principal components with both types of random effects the DIC was 8674, which suggested a better model fit than the model without spatially structured random effects (DIC=8784). However, in the model including a spatial random effect, the pD was negative (pD= -15.42).

The model examining associations between aesthetic objectives PCs and diabetes incidence with the spatial random effects included had a DIC of 5868, while the model in which the spatially structured random effects were excluded had a DIC of 5894.

#### *5.4.9 Prior sensitivity*

Sensitivity to different specifications of the hyperpriors for precision was evaluated on the model for diabetes incidence for the private water supply data, with arsenic concentrations categorized by quintile. The coefficients for the various arsenic concentrations were similar regardless of the hyperprior specification (Table 5.22).

**Table 5.22** Results of prior sensitivity analysis comparing the reported Bayesian model of associations between arsenic concentrations in private wells categorized by quintiles and diabetes incidence in 2010 to models with two alternative hyperprior specifications for the unstructured and structured random effects.

Effect	Reported model			Alternative hyperpriors on random effects					
	$\sim \text{gamma}(0.001, 0.001)$			$\sim \text{gamma}(0.5, 0.0005)$			$\text{precision} = 1/SD^2 \text{ where } SD \sim \text{uniform}(0, 5)$		
	95% CrI			95% CrI			95% CrI		
	RR	lower	upper	RR	lower	upper	RR	lower	upper
Intercept	1.054	0.955	1.161	1.054	0.958	1.159	1.052	0.952	1.161
Arsenic Q1	reference category			reference category			reference category		
Arsenic Q2	0.933	0.833	1.046	0.931	0.834	1.040	0.934	0.833	1.050
Arsenic Q3	0.999	0.895	1.119	0.996	0.895	1.108	1.001	0.896	1.121
Arsenic Q4	0.937	0.840	1.047	0.936	0.843	1.042	0.938	0.840	1.051
Arsenic Q5	0.854	0.761	0.958	0.856	0.767	0.956	0.855	0.760	0.960
COPD	1.816	1.055	3.086	1.847	1.076	3.133	1.803	1.045	3.047
Education <sup>1</sup>	0.879	0.736	1.049	0.880	0.728	1.050	0.880	0.738	1.051
Income <sup>2</sup>	0.998	0.995	1.001	0.998	0.995	1.001	0.998	0.995	1.002
Random Effects		SD	SE		SD	SE		SD	SE
Unstructured RE		0.077	0.029		0.072	0.033		0.076	0.035
Spatially structured RE		0.070	0.028		0.050	0.032		0.073	0.032

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts, and by prevalence of COPD, education level and income by inclusion as risk factors in model.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total annual income (thousands of dollars).

CrI = credible interval, CI = confidence interval, RR = risk ratio, PC = principal component, SD = standard deviation SE = standard error, RE = random effect.

## 5.5 Discussion

There was no evidence that groundwater arsenic concentrations in public supplies or private wells were associated with an increased risk of diabetes in our study area. A significant, but non-linear and not clearly monotonic, association was identified between increasing arsenic concentrations in water from private wells and decreasing diabetes incidence in 2010-2012. This protective effect was in contrast to what was expected based on previous studies (Maull et al., 2012; Kuo et al., 2013; Wang et al., 2014) and was only significant for the highest quintile compared to the lowest quintile. No significant associations between drinking water arsenic concentrations and diabetes incidence were evident in frequentist GLMMs for the four previous time periods, suggesting that the association between arsenic concentration and diabetes incidence for 2010 – 2012 was very likely a spurious finding. Increasing arsenic concentrations

in groundwater were not associated with an increase in diabetes incidence or prevalence in Saskatchewan.

When examining the health standards PC scores in public water supplies, there was an association between increased scores for the first health standard PC and increased diabetes prevalence. The first PC for public supplies was characterized mainly by nitrate and selenium, with a moderate loading of lead. However, it appeared that the form of this relationship was not linear and also not strictly monotonic, particularly with respect to the effect size being slightly but not significantly greater for the third than for the fourth quintile.

The association between PC1<sub>health</sub> and diabetes prevalence was not as apparent in the frequentist model for 2010, suggesting that unmeasured, spatially related confounding variables could be masking the association in frequentist models. Comparing the DIC for the model for diabetes prevalence with health standards principal components as risk with and without spatially structured random effects suggested a superior model fit when the spatially correlated random effects were included in the model. This was consistent with the identification of global spatial autocorrelation in the residuals from the frequentist model with no spatial structure. However, the pD was negative for the model with spatial random effects. A negative pD renders the DIC uninterpretable (Spiegelhalter et al., 2002) and could indicate a problem with model configuration or convergence. In addition, the Kaiser-Meyer-Olkin measure of sampling adequacy suggested that overall the amount of shared variance in the health standards from public supplies was low (Chapter 4). Therefore, PCA may not have adequately summarized the variance for health standards in public supplies. Consequently, this model and its findings should be interpreted with caution. The finding that increased PC1 scores were associated with increased risk of diabetes is consistent with recent evidence suggesting that high levels of dietary selenium are associated with the prevalence of diabetes (Stranges et al., 2007; Wei et al., 2015; Lu et al., 2016). However, selenium exceeded the provincial drinking water standard of 0.01 mg/L in less than 2% of samples and was below detection limits in 72% of samples from public water supplies (Chapter 4). In combination with the caution warranted in the interpretation of the spatial model, a lack of significant associations in the frequentist models for other years, and a high probability of falsely identifying an association by chance among all the variables assessed

for associations between water quality and diabetes prevalence in 2010, it is probable that this result was a spurious finding.

The apparent protective associations in the frequentist analysis between increasing PC2<sub>health</sub> scores for public water supplies and decreasing diabetes prevalence in 4 of the 5 years was not significant after accounting for spatial autocorrelation using Bayesian analysis for the 2010 data. It is possible that unmeasured, spatially correlated confounders accounted for the apparent association, although caution is warranted in interpretation of the spatial model as explained previously. PC2<sub>health</sub> in public supplies was characterized by strong positive loading of boron and strong negative loading of barium; a plausible explanation for the relationship of these elements with diabetes could not be identified in a literature search.

Although we hypothesized that unpalatable drinking water could drive consumption of sugar sweetened beverages and indirectly contribute to the development of diabetes, no clear associations were found between living in areas with drinking water of poor aesthetic quality and an increase in the incidence or prevalence of diabetes. An apparent protective association between PC3<sub>aesthetic</sub> for private wells and diabetes incidence in the Bayesian model for 2010-2012 was not evident when the scores were categorized into quintiles. This protective association was not evident in any of the frequentist GLMMs for private well data, nor any of the models using public water supply data.

The lack of consistent findings between time periods suggests the isolated significant result for the private well data may have been a chance finding, especially since no plausible explanation could be found for a protective effect against diabetes by the components of PC3<sub>aesthetic</sub>, characterized primarily by the presence of iron and manganese.

Examination of different model specifications and various time periods permitted a weight of evidence approach in assessing potential associations between diabetes and various aspects of water quality. However, examination of multiple years also increased the chance of finding a statistically significant association where none exists due to making multiple comparisons. The consistency of results across the two types of water supplies and five time periods were considered, along with the risk of finding a false association by chance due to multiple testing. Putative associations were also assessed for plausibility and consistency with previously published literature examining associations between water quality and diabetes. The present

study identified no substantial evidence for causal associations between arsenic concentration, health standards PC scores, or aesthetic objectives PC scores and diabetes incidence or prevalence.

Bayesian disease modeling techniques were used to incorporate neighborhood structure into models and recognize the potential for unmeasured risk factors that varied based on location in the province. The Bayesian models do not just account for clustering of outcomes within regions, but adjust for spatial correlation between neighboring regions. Estimates of disease risk are smoothed as a result of borrowing information from neighboring areas. This is especially important for area data when case counts are expected to be small and effect estimates subject to random fluctuation over time. Comparison of model fit with and without the spatial random effect using the deviance information criterion (DIC) indicated that including the spatial random effect improved the fit of models of diabetes incidence and prevalence for the models where potential associations were identified. The incorporation of both a structured spatial random effect and an unstructured random effect also allows partitioning of variance unexplained by the model into variance that is related to spatial location and variance that represents random noise, which can inform future studies investigating risk factors for disease. The statistically significant tests for global autocorrelation for the area-level residuals from the majority of the frequentist GLMM outcome and risk factor combinations further supported the incorporations of spatially correlated random effects into models of risk factors for diabetes in rural Saskatchewan.

Hierarchical random effect models run in a frequentist framework were used for comparison to the Bayesian models run in OpenBUGS. Because these models were computationally more efficient, this allowed comparison of multiple years of data to assess the weight of evidence for any associations found. The effects of the water variables were often inconsistent between years and also between public water supply and private well data where the variables were directly comparable. Conversely, the estimates for the other covariates included in the model as possible confounding variables (COPD prevalence, education and income) were very consistent when compared between models for the different types of water supplies, and appeared to be relatively consistent across time periods. The variability in the results for the water variables across different time periods and outcomes demonstrates that spurious findings may hamper the ability

to accurately determine relationships between water quality and diabetes rates in some studies depending on the time period and case definitions used.

Not completing high school had an unexpected protective association with diabetes, although the association was not statistically significant. A recent study found that not completing high school, compared to completing some post secondary education other than university, was associated with increased rates of diabetes in rural Saskatchewan residents, although after controlling for other factors, education was not a significant predictor of diabetes in the final model (Dyck et al., 2013). The unexpected result in the present study could reflect that completing or not completing high school may not be a sensitive measure of the level of education that could influence diabetes rates in rural Saskatchewan. The finding could also have been an artefact of the ecological nature of the education covariate.

This study suffers from the same limitations as other ecological analyses, particularly the problem of the ecological fallacy. We assessed diabetes cases and water exposure at a group level, but do not know the water exposure of the diabetes cases as compared to non-cases. In this case we did not find any robust associations between water quality and diabetes but it is possible that the generalization of diabetes rates and exposures over regions may have masked small effects that could be evident if studied at an individual level.

Summarizing exposures over geographic regions was further complicated by the fact that residents could have accessed a private water source, a public water supply, or a combination of these, or neither if they do not consume tap water. In addition, the type of water supply used could also have varied over time. Furthermore, the content of the household water can be altered significantly by in-home water treatment systems before consumption. In response to a questionnaire administered to rural Saskatchewan residents about tap water consumption, 47.6% of respondents reported treating their household tap water (McLeod et al., 2014). Along with water treatment implementation, individual consumption could also be influenced by perceptions of safety and quality (McLeod et al., 2014). Therefore, individual consumption patterns and cumulative exposure could vary considerably between residents within a region as well as among those consuming an identical water supply.

Our exposure assessment was also subject to potential misclassification due to the challenge of estimating exposure over large geographic regions. We used geostatistical methods to determine

a mean value for arsenic concentrations and principal components scores for the regions in our study.

Kriging was previously validated as a method useful in assessing exposure to inorganic arsenic from drinking water from wells (Meliker et al., 2008; James et al., 2014) and represents a valuable tool for estimating exposures over large regions. Although wells with moderately elevated arsenic concentrations were present in our study area and exhibited some clustering, there was much variation in the arsenic concentrations of neighboring wells. However, most public supplies and private wells fell below the drinking water standard of 10 µg/L for arsenic concentration, and a high proportion had arsenic concentrations that were not detectable by analytic methods available (Chapter 4). Ultimately, the concentrations predicted by kriging rarely exceeded the drinking water standard, and the averaging of interpolated values over relatively large regions meant that the concentrations of arsenic used as measures of arsenic exposure were well below the standard of 10 µg/L and lower than the concentrations analyzed in many other studies (Navas-Acien et al., 2005; Maull et al., 2012; Kuo et al., 2013; Wang et al., 2014). This would most likely have biased our results towards the null, given that recent reviews concluded that there is insufficient evidence to conclude that low to moderate arsenic concentrations are associated with diabetes (Maull et al., 2012; Kuo et al., 2013).

The use of PCA combined with kriging also resulted in a narrow range of PC scores with little variability between regions. PCA is a statistical technique that is used to reduce the dimensionality of a data set by grouping correlated variables based on their covariance. Kriging of PC scores has not been previously validated for prediction of PC scores despite being relatively commonly used to characterize groundwater constituents over larger regions (Rao et al., 2009; Shyu et al., 2011). Prior to PCA and kriging, the public supply data were also summarized by the use of general linear mixed models with an exponential correlation structure to account for repeated measures at each sampling location, which decreased the variability in the data with generalization of these values towards the overall mean. Summarizing groundwater variables with this combination of techniques represents a source of potential misclassification and would likely bias results toward the null, due to moderation of more extreme values and generalization toward the overall mean values for all regions.

Accurate residential information for rural residents is difficult to obtain from the population-based administrative data used in this study. We used rural municipality portion of the residence code to place residents in the geographic units used for analysis. However, this code is assigned to residents based on a hierarchy of residence information starting with land location followed by mailing address. In rural areas where mail delivery is to a regional post office rather than residential, it is possible that the mailing address could be from a different rural municipality than the place of residence, depending on the nearest post office location (eHealth Saskatchewan, 2015). It is not known what proportion of residence information in the administrative health data was based on mailing address or the rate of correspondence of mailing address RM to the actual residential RM; it is likely that there were some cases of misclassification of residents to geographic units.

This issue was also likely exacerbated for those registered to a First Nation; historically the residence code was based on band affiliation so could place residents on a home reserve regardless of actual place of residence. This system under revision to base residence on postal code and while corrections have been made, there is uncertainty of the accuracy of residence information First Nations residents for our cohorts. Residents identified as First Nations were limited to those who with Registered Indian status, which excludes residents with First Nations heritage but who did not apply or qualify for inclusion in the registry. For this reason, it is likely we underestimated the population of First Nations persons in our analysis.

We also made the assumption that residential mobility was minimal in our study population. Place of residence was established at the beginning of each time period, with the assumption that residents were exposed to the water at that place of residence for some time prior to their identification as an incident or prevalent diabetes case. Overall, the study population tended to be non-mobile, given that 76.5% of our study population were placed in the same RM for all years from 2004–2012. However, the assumption that place of residence reflects historical exposure to the water in that location does mean our exposure assessment was imprecise in that we do not know how long study subjects were exposed to the water quality variables being assessed in this study and some exposures may have been misclassified.

Exposure assessment is a considerable challenge for studies of associations between drinking water and diabetes, even for individual-level studies. Determining historical arsenic



concentrations in water supplies along with an accurate consumption history is very difficult. A poor understanding of the mechanisms by which arsenic in drinking water might influence the development of diabetes further complicates accurate exposure assessment over the relevant induction period. As Rothman (Rothman, 1981) suggests, studies that do not accurately account for the empirical induction period represent a form of non-differential misclassification of exposure that results in bias toward the null that can underestimate or mask real effects, and lead to conflicting results. Given that we had no means to accurately assess the time frame of exposure to the risk factors we evaluated, this form of bias was likely a factor in our study, and may have masked any associations between our risk factors and diabetes.

While our objective was to investigate associations between type 2 diabetes and water quality, our case definition would have identified physician visits for all diabetes cases. This was because the International Classification of Diseases 9<sup>th</sup> Edition (ICD-9) codes were used to identify cases and a type 2 diabetes-specific code was not available in ICD-9. The exclusion of individuals under 35 years of age minimized the likelihood of including newly identified type 1 diabetes in our incident cases because most cases of type 1 diabetes occur in children and youth (Public Health Agency of Canada, 2011). However, our analysis of diabetes prevalence would have included cases of type 1 diabetes, although it is estimated that 90-95% of cases of diabetes in Canada are type 2 diabetes (Public Health Agency of Canada, 2011).

In addition, the use of ICD codes to identify diabetes cases could result in misclassification of incident and prevalent diabetes cases. While case identification algorithms have been validated, the reported sensitivity of the algorithm used for identifying diabetes cases was 92.3% and the specificity was 96.9% (Chen et al., 2010). Furthermore, in 2007, chronic disease management codes were implemented for physician fee claims so that codes could be included for all chronic diseases under treatment. Prior to 2007 physician claims included only one diagnosis per claim; if a patient with diabetes visited a physician but diabetes was not identified as the primary reason the visit, diabetes would not have been included in the claim. Consequently, the number of physician visits for diabetes could have been underestimated prior to 2007.

Our analysis was missing important known risk factors for diabetes such as body mass index and level of physical activity (Public Health Agency of Canada, 2011; World Health Organization, 2016). Data on these risk factors, while available at a coarse spatial resolution in the Canadian

Community Health Survey (by Statistics Canada), were not available at across our rural study area at a resolution that would have contributed useful information to our analysis. These risk factors would likely be important in explaining some of the residual variance in our models, and would be important to consider in future studies of diabetes in Saskatchewan.

Data regarding smoking rates in rural Saskatchewan were also not available at a resolution congruous with the geographic areas used in our analysis. Prevalence of COPD was used as a proxy for smoking rates, given its strong association with COPD (Single et al., 2000) and the availability of stratum-specific COPD prevalence based on administrative health data.

The education and income covariate data was obtained from Statistics Canada and was matched to the stratification for diabetes outcomes as closely as possible. These data were available at the level of census subdivision, an administrative unit smaller than the geographic areas used for the analysis, and were summed to obtain estimates for the analysis. Census Subdivisions sometimes represent very small populations in rural Saskatchewan. Suppression of these data for confidentiality reasons, necessitated limited imputation of missing values. Combined with the random rounding applied to educational attainment data, the socioeconomic covariate data could be subject to error, especially in areas with small populations.

While acknowledging limitations in our data, we were able to use existing administrative and surveillance data for both diabetes incidence and prevalence and water quality to analyze potential spatial associations between numerous aspects of water quality and diabetes and screen for any large scale effects. Although some variables were available only on an ecological level, available individual level data was incorporated into the analysis through stratification and indirect standardization for sex, age and First Nations status, providing advantages over a strictly ecological analysis. By using hierarchical Bayesian modeling with the incorporation of both unstructured and spatially correlated random effects and comparing this to frequentist GLMM models for multiple time periods we were able to use a weight of evidence approach to examine any potential associations that were identified, while carefully considering the limitations in the types of data that were available to us.

## 5.6 Conclusions

No clear associations between groundwater quality and diabetes incidence or prevalence were apparent from the data available for this study. A detailed comparison of Bayesian multilevel models and frequentist GLMMs for multiple time periods revealed no compelling evidence for any associations between arsenic or principal components summarizing health standards and aesthetic objectives from the Saskatchewan Drinking Water Quality Standards and Objectives and increased risk of diabetes. Because this was an ecological study, we cannot rule out the presence of local or individual-level effects of arsenic or other health- or aesthetic-related aspects of water quality in groundwater drinking supplies on diabetes. As groundwater quality can be highly variable over small distances, exposure assessment at an individual level, although challenging, would be necessary to better evaluate the effects of water quality on diabetes incidence and prevalence. However, based on the results of our analysis, there was no evidence for large scale area-level effects of water quality on diabetes risk in Saskatchewan.

## 5.7 References

- Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1–20. doi:10.1007/BF00116466
- Chen, G., Khan, N., Walker, R., Quan, H., 2010. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Research and Clinical Practice* 89, 189–195. doi:10.1016/j.diabres.2010.03.007
- Chen, Y., Ahsan, H., Slavkovich, V., Peltier, G.L., Gluskin, R.T., Parvez, F., Liu, X., Graziano, J.H., 2010. No association between arsenic exposure from drinking water and diabetes mellitus: a cross-sectional study in Bangladesh. *Environmental health perspectives* 118, 1299–1305.
- de Koning, L., Malik, V.S., Rimm, E.B., Willett, W.C., Hu, F.B., 2011. Sugar-sweetened and artificially sweetened beverage consumption and risk of type 2 diabetes in men. *Am J Clin Nutr* 93, 1321–1327. doi:10.3945/ajcn.110.007922
- de Vocht, F., Hannam, K., Buchan, I., 2013. Environmental risk factors for cancers of the brain and nervous system: the use of ecological data to generate hypotheses. *Occup Environ Med* 70, 349–356. doi:10.1136/oemed-2012-100954
- Dyck, R., Karunanayake, C., Pahwa, P., Hagel, L., Lawson, J., Rennie, D., Dosman, J., 2013. Prevalence, risk factors and co-morbidities of diabetes among adults in rural Saskatchewan: the influence of farm residence and agriculture-related exposures. *BMC Public Health* 13, 7. doi:10.1186/1471-2458-13-7
- Dyck, R., Osgood, N., Lin, T.H., Gao, A., Stang, M.R., 2010. Epidemiology of diabetes mellitus among First Nations and non-First Nations adults. *Canadian Medical Association Journal* 182, 249–256. doi:10.1503/cmaj.090846
- eHealth Saskatchewan, 2015. Covered Population 2015: Notice to Readers. URL <http://population.health.gov.sk.ca/NoticetoReaders2015.htm>.
- Elliott, P., Savitz, D.A., 2008. Design Issues in Small-Area Studies of Environment and Health. *Environ Health Perspect* 116, 1098–1104. doi:10.1289/ehp.10817
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical science* 7, 457–472.
- Gershon, A.S., Wang, C., Guan, J., Vasilevska-Ristovska, J., Cicutto, L., To, T., 2009. Identifying Individuals with Physician Diagnosed COPD in Health Administrative Databases. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 6, 388–394. doi:10.1080/15412550903140865
- Government of Saskatchewan, 2002. Drinking Water - Source to Tap Solutions.
- Hux, J.E., Ivis, F., Flintoft, V., Bica, A., 2002. Diabetes in Ontario Determination of prevalence and incidence using a validated administrative data algorithm. *Dia Care* 25, 512–516. doi:10.2337/diacare.25.3.512
- James, K.A., Meliker, J.R., Battenfield, B.E., Byers, T., Zerbe, G.O., Hokanson, J.E., Marshall, J.A., 2014. Predicting arsenic concentrations in groundwater of San Luis Valley,

- Colorado: implications for individual-level lifetime exposure assessment. *Environmental geochemistry and health* 36, 773–782.
- Jang, M.J., Lee, Y., Lawson, A.B., Browne, W.J., 2007. A comparison of the hierarchical likelihood and Bayesian approaches to spatial epidemiological modelling. *Environmetrics* 18, 809–821.
- Künzli, N., Tager, I.B., 1997. The semi-individual study in air pollution epidemiology: a valid design as compared to ecologic studies. *Environ Health Perspect* 105, 1078–1083.
- Kuo, C.-C., Moon, K., Thayer, K.A., Navas-Acien, A., 2013. Environmental chemicals and type 2 diabetes: an updated systematic review of the epidemiologic evidence. *Current diabetes reports* 13, 831–849.
- Lawson, A.B., 2013. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, Second Edition. Chapman and Hall/CRC, Boca Raton.
- Li, X., Li, B., Xi, S., Zheng, Q., Lv, X., Sun, G., 2013. Prolonged environmental exposure of arsenic through drinking water on the risk of hypertension and type 2 diabetes. *Environ Sci Pollut Res Int* 20, 8151–8161. doi:10.1007/s11356-013-1768-9
- Lu, C.-W., Chang, H.-H., Yang, K.-C., Kuo, C.-S., Lee, L.-T., Huang, K.-C., 2016. High serum selenium levels are associated with increased risk for diabetes mellitus independent of central obesity and insulin resistance. *BMJ Open Diab Res Care* 4, e000253. doi:10.1136/bmjdr-2016-000253
- Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique and future directions. *Statist. Med.* 28, 3049–3067. doi:10.1002/sim.3680
- Malik, V.S., Popkin, B.M., Bray, G.A., Després, J.-P., Hu, F.B., 2010. Sugar-Sweetened Beverages, Obesity, Type 2 Diabetes Mellitus, and Cardiovascular Disease Risk. *Circulation* 121, 1356–1364. doi:10.1161/CIRCULATIONAHA.109.876185
- Maull, E.A., Ahsan, H., Edwards, J., Longnecker, M.P., Navas-Acien, A., Pi, J., Silbergeld, E.K., Styblo, M., Tseng, C.-H., Thayer, K.A., Loomis, D., 2012. Evaluation of the Association between Arsenic and Diabetes: A National Toxicology Program Workshop Review. *Environmental Health Perspectives*. 120, 1658–1670. doi:10.1289/ehp.1104579
- McLeod, L., Bharadwaj, L., Waldner, C., 2015. Risk factors associated with perceptions of drinking water quality in rural Saskatchewan. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques* 40, 36–46. doi:10.1080/07011784.2014.985513
- McLeod, L., Bharadwaj, L., Waldner, C., 2014. Risk Factors Associated with the Choice to Drink Bottled Water and Tap Water in Rural Saskatchewan. *International Journal of Environmental Research and Public Health* 11, 1626–1646. doi:10.3390/ijerph110201626
- Meliker, J.R., AvRuskin, G.A., Slotnick, M.J., Goovaerts, P., Schottenfeld, D., Jacquez, G.M., Nriagu, J.O., 2008. Validity of spatial models of arsenic concentrations in private well water. *Environmental Research* 106, 42–50. doi:10.1016/j.envres.2007.09.001
- Navas-Acien, A., Silbergeld, E.K., Pastor-Barriuso, R., Guallar, E., 2008. Arsenic exposure and prevalence of type 2 diabetes in US adults. *JAMA: the journal of the American Medical Association* 300, 814–822.

- Navas-Acien, A., Silbergeld, E.K., Streeter, R.A., Clark, J.M., Burke, T.A., Guallar, E., 2006. Arsenic Exposure and Type 2 Diabetes: A Systematic Review of the Experimental and Epidemiologic Evidence. *Environmental Health Perspectives* 114, 641–648. doi:10.1289/ehp.8551
- Norman, R.E., Carpenter, D.O., Scott, J., Brune, M.N., Sly, P.D., 2013. Environmental exposures: an underrecognized contribution to noncommunicable diseases. *Reviews on Environmental Health* 28, 59–65. doi:10.1515/reveh-2012-0033
- Onufrak, S.J., Park, S., Sharkey, J.R., Sherry, B., 2014. The relationship of perceptions of tap water safety with intake of sugar-sweetened beverages and plain water among US adults. *Public Health Nutrition* 17, 179–185. doi:10.1017/S1368980012004600
- Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R news* 6, 7–11.
- Prüss-Ustün, A., Vickers, C., Haeffliger, P., Bertollini, R., 2011. Knowns and unknowns on burden of disease due to chemicals: a systematic review. *Environmental Health* 10, 9. doi:10.1186/1476-069X-10-9
- Public Health Agency of Canada, 2011. Diabetes in Canada: Facts and figures from a public health perspective. URL <http://www.phac-aspc.gc.ca/cd-mc/publications/diabetes-diabete/facts-figures-faits-chiffres-2011/index-eng.php>.
- Public Health Agency of Canada, 2009. National Diabetes Surveillance System Methods Documentation, 2008.
- Rothman, K.J., 1981. Induction and Latent Periods. *Am. J. Epidemiol.* 114, 253–259.
- Rothman, K.J., Greenland, S., Lash, T.L., 2008. *Modern Epidemiology*. Lippincott Williams & Wilkins, Philadelphia, PA.
- Satyaji Rao, Y.R., Keshari, A.K., Gosain, A.K., 2009. Evaluation of regional groundwater quality using PCA and geostatistics in the urban coastal aquifer, East Coast of India. *International Journal of Environment and Waste Management* 5, 163–180.
- Shyu, G.-S., Cheng, B.-Y., Chiang, C.-T., Yao, P.-H., Chang, T.-K., 2011. Applying Factor Analysis Combined with Kriging and Information Entropy Theory for Mapping and Evaluating the Stability of Groundwater Quality Variation in Taiwan. *Int J Environ Res Public Health* 8, 1084–1109. doi:10.3390/ijerph8041084
- Single, E., Rehm, J., Robson, L., Truong, M.V., 2000. The relative risks and etiologic fractions of different causes of death and disease attributable to alcohol, tobacco and illicit drug use in Canada. *CMAJ* 162, 1669–1675.
- Sketchell, J., Shaheen, N., 2000. Ground water quality in rural Saskatchewan—Emerging issues for drinking water, in: *Maintaining Drinking Water Quality—Lessons from the Prairies and Beyond*. Proceedings of the 9th National Conference on Drinking Water. Regina. Saskatchewan. Canada. (Ed. W. Robertson.). pp. 242–258.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 583–639.

- Statistics Canada, 2008a. 2006 Census: Data quality and confidentiality standards and guidelines (public): Data quality practices. URL <https://www12.statcan.gc.ca/census-recensement/2006/ref/notes/DQ-QD/DQPractices-QDPratiques-eng.cfm>.
- Statistics Canada, 2008b. Data quality and confidentiality standards and guidelines (public): Confidentiality (non-disclosure) rules. URL <https://www12.statcan.gc.ca/census-recensement/2006/ref/notes/DQ-QD/confidentiality-confidentialite-eng.cfm>.
- Steinmaus, C., Yuan, Y., Liaw, J., Smith, A.H., 2009. Low-level Population Exposure to Inorganic Arsenic in the United States and Diabetes Mellitus. *Epidemiology* 20, 807–815. doi:10.1097/EDE.0b013e3181b0fd29
- Stranges, S., Marshall, J.R., Natarajan, R., Donahue, R.P., Trevisan, M., Combs, G.F., Cappuccio, F.P., Ceriello, A., Reid, M.E., 2007. Effects of long-term selenium supplementation on the incidence of type 2 diabetes: a randomized trial. *Ann. Intern. Med.* 147, 217–223.
- Thayer, K.A., Heindel, J.J., Bucher, J.R., Gallo, M.A., 2012. Role of environmental chemicals in diabetes and obesity: a National Toxicology Program workshop review. *Environmental Health Perspectives* 120, 779–789.
- Thompson, T.S., 2003. General Chemical Water Quality of Private Groundwater Supplies in Saskatchewan, Canada. *Bulletin of Environmental Contamination and Toxicology* 70, 0447–0454. doi:10.1007/s00128-003-0007-3
- Thompson, T.S., 2001. Nitrate concentrations in private rural drinking water supplies in Saskatchewan, Canada. *Bulletin of environmental contamination and toxicology* 66, 64–70.
- Thompson, T.S., Le, M.D., Kasick, A.R., Macaulay, T.J., 1999. Arsenic in Well Water Supplies in Saskatchewan. *Bulletin of Environmental Contamination and Toxicology* 63, 478–483. doi:10.1007/s001289901005
- Wang, W., Xie, Z., Lin, Y., Zhang, D., 2014. Association of inorganic arsenic exposure with type 2 diabetes mellitus: a meta-analysis. *J Epidemiol Community Health* 68, 176–184. doi:10.1136/jech-2013-203114
- Water Security Agency, n.d. Saskatchewan's Drinking Water Quality Standards and Objectives (Summarized). URL <http://www.saskh2o.ca/pdf/epb507.pdf>
- Wei, J., Zeng, C., Gong, Q., Yang, H., Li, X., Lei, G., Yang, T., 2015. The association between dietary selenium intake and diabetes: a cross-sectional study among middle-aged and older adults. *Nutrition Journal* 14, 18. doi:10.1186/s12937-015-0007-2
- World Health Organization, 2016. *Global Report on Diabetes*. Geneva, Switzerland.
- Zierold, K.M., Knobeloch, L., Anderson, H., 2004. Prevalence of Chronic Diseases in Adults Exposed to Arsenic-Contaminated Drinking Water. *Am J Public Health* 94, 1936–1937.

## **CHAPTER 6: ECOLOGICAL ANALYSIS OF ASSOCIATIONS BETWEEN GROUNDWATER QUALITY AND HYPERTENSION AND CARDIOVASCULAR DISEASE IN RURAL SASKATCHEWAN USING BAYESIAN HIERARCHICAL MODELS**

**Disclaimer: this study is based in part on de-identified data provided by the Saskatchewan Ministry of Health. The interpretation and conclusions contained herein do not necessarily represent those of the Government of Saskatchewan or the Ministry of Health.**

*The research in this chapter uses the summarized water quality data from Chapter 4 to evaluate associations between water quality and cardiovascular disease, specifically hypertension, ischemic heart disease and stroke. As for the diabetes outcomes, Bayesian hierarchical models that account for spatial relationships in the outcomes and risk factors were used. Direct and indirect effects of mixtures of water contaminants, assessed through the use of principal components, as well as arsenic concentrations were investigated for associations with cardiovascular disease. The application of innovative techniques, using existing data sources, can be a powerful tool to identify aspects of water quality that are of concern, informing future research needs and priorities for public health efforts to ensure water quality is not adversely impacting the health of Saskatchewan's rural population.*



## 6.1 Abstract

Associations between groundwater quality and the prevalence of hypertension, ischemic heart disease, and stroke were investigated using an ecological study design and Bayesian spatial analysis. Previously collected water quality surveillance data from public water supplies and private wells were accessed to estimate exposures, and administrative health data were accessed to estimate health outcomes. Water quality exposures were estimated by applying geostatistical techniques to arsenic concentrations and principal component scores used to summarize groups of parameters measured as either health standards or aesthetic objectives described in Saskatchewan's Drinking Water Quality Standards and Objectives (Water Security Agency, n.d.). Generalized linear mixed models with a log link were used to assess associations between the water quality variables and health outcomes. Bayesian hierarchical models for prevalence of each outcome in 2010 were considered the primary evidence for associations and compared to frequentist models for 2006-2010 with no spatial random effects. Effect estimates were controlled for sex and age by stratification of case counts and expected case counts, for smoking by inclusion of sex- and age-specific prevalence of chronic obstructive pulmonary disease as a surrogate covariate, and for education and income by use of stratified variables derived from census data.

There was no evidence for associations between groundwater arsenic concentrations in public or private water supplies and increased risk of hypertension or cardiovascular disease. Among the Bayesian models for 2010, an association was identified between increased scores for the first principal component for health standards from public water supplies and increased stroke prevalence. In addition, the second aesthetic objectives principal component scores from public supplies demonstrated a protective effect against ischemic heart disease, effects consistent with previous literature. Similarly, the second aesthetic objectives principal component in private supplies was associated with decreased prevalence of hypertension. In public supplies, the third aesthetic objective principal component was associated with decreased prevalence of stroke. In the frequentist models for 2006-2010 several protective associations were consistently demonstrated between various principal component scores and ischemic heart disease and hypertension as well as between arsenic concentrations in public supplies and hypertension. However, these protective relationships were not apparent in the Bayesian models, which

suggested that the associations evident in frequentist models were potentially due to unmeasured confounders.

The results of this study illustrate the complexity of relationships between drinking water exposures and hypertension and cardiovascular disease. Further investigation is warranted at the individual level, particularly with respect to the potential beneficial effect of hard water on the prevalence of hypertension and cardiovascular disease.

## **6.2 Introduction**

Heart disease and stroke were the second and third leading causes of death in Canada in 2012 (Public Health Agency of Canada, 2015) and are leading contributors to the economic burden of disease in Canada (Public Health Agency of Canada, 2014). Hypertension, a major risk factor for cardiovascular diseases, affected 19.6% of the population in Canada in 2007/2008 (Robataille et al., 2012) and has been described as a global public health crisis (WHO 2013). While a variety of lifestyle risk factors have been established for hypertension and cardiovascular disease (Yusuf et al., 2001; World Health Organization, 2003), evidence also suggests that exposure to environmental pollutants contributes to the development of cardiovascular disease (Bhatnagar, 2006). Exposure to arsenic in drinking water has been linked to hypertension (Abir et al., 2011; Abhyankar et al., 2012) as well as ischemic heart disease (IHD) and stroke (Navas-Acien et al., 2005; Wang et al., 2007; Moon et al., 2012; Tsuji et al., 2014). While associations between arsenic and hypertension and cardiovascular disease have been demonstrated in areas where arsenic concentrations in groundwater were  $> 500 \mu\text{g/L}$  (Chen et al., 1995; Chen et al., 1996; Rahman et al., 1999; Tseng et al., 2003; Wang et al., 2007), there is growing concern that low to moderate ( $<100 \mu\text{g/L}$ ) drinking water arsenic concentrations may also be associated with these diseases (Gong and O'Bryant, 2012; Moon et al., 2012; James et al., 2015).

Exposure to arsenic in drinking water could represent an important modifiable risk factor that could help mitigate the burden of hypertension and cardiovascular disease in impacted regions (Wang et al., 2011; Abhyankar et al., 2012). Arsenic is a toxic metalloid ubiquitous in the environment from both natural and anthropogenic sources. Dissolution from mineral deposits can result in naturally occurring arsenic contamination of groundwater. Arsenic concentration in

groundwater can be very high where arsenic occurs naturally in bedrock (Thompson et al., 1999), and drinking water is considered a major route of arsenic exposure globally (Flora, 2014).

Other attributes of tap water quality in addition to the presence of natural or anthropogenic toxins can also affect health. Poor tap water quality can discourage the consumption of tap water, instead favoring the consumption of sugar sweetened beverages (Onufrak et al., 2014), which in turn is a risk factor for weight gain, and may increase the risk of cardiovascular disease independent of the effects of increased body weight (Malik et al., 2010). Areas with poor quality drinking water could consequently experience an increase in the prevalence of hypertension and cardiovascular disease mediated by indirect effects of poor quality drinking water especially where the palatability of drinking water is affected. Tap water of poor quality could also potentially contribute to the development of hypertension (Tuthill and Calabrese, 1979) either directly through the sodium content of water or through the consumption of softened water (Padwal et al., 2005). Though the relationship between salt intake and hypertension is uncertain and complex, reduced salt consumption continues to be recommended for prevention and disease management (Frisoli et al., 2012).

The primary sources for drinking water and extent of testing varies depending on area of residence in Canada. Approximately 43% of Saskatchewan residents rely on groundwater for domestic use, primarily in rural areas and smaller municipalities (Environment and Climate Change Canada, 2007) and 14% obtain household water from private wells (Government of Saskatchewan, 2002). As is the case in many jurisdictions, no regulations govern the safety of private water supplies in Saskatchewan. Owners of private water supplies have sole responsibility for testing and treatment of their water supplies, including associated costs. While provincial agencies do monitor the quality of public drinking water, the requirements for public supplies vary depending on the size of the population served by the supply and the type of water source. Consequently, public water supplies for smaller communities are typically not monitored at an intensity comparable to water supplies in cities. Differential testing and regulation of water supplies for residents of rural areas could increase the likelihood of exposure to poor quality drinking water.

Provinces in Canada establish their own regulations pertaining to drinking water. Saskatchewan established two main categories of drinking water guidelines: legally enforceable standards that

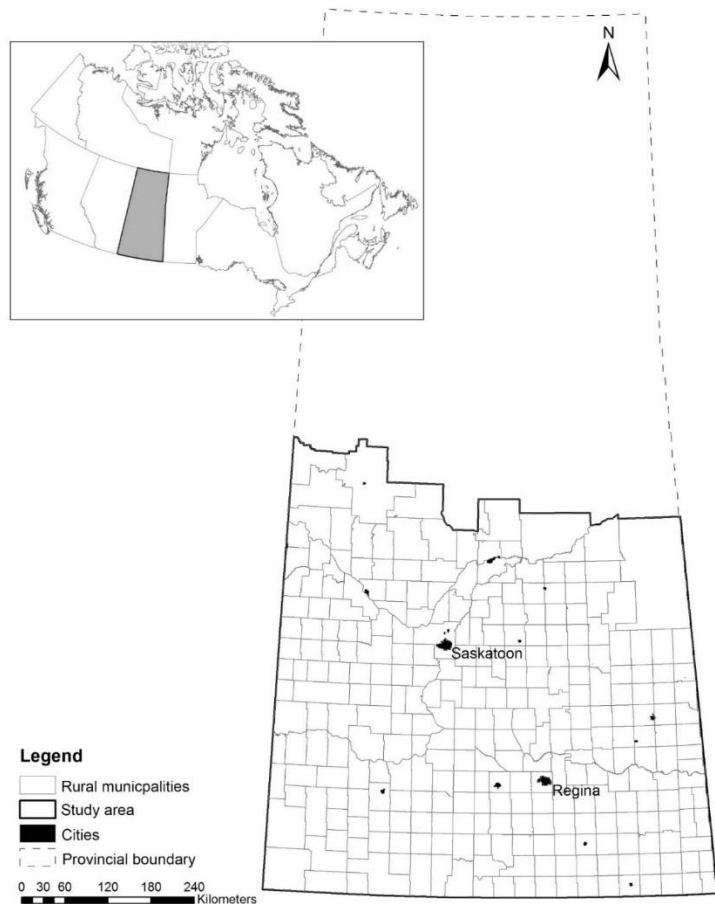
govern health hazards (heavy metals, pesticides, microorganisms), and objectives that represent optimal levels of characteristics which primarily affect the aesthetic qualities of drinking water but are not health hazards (Water Security Agency, n.d.). Previous studies have identified wells with elevated concentrations of arsenic (Thompson et al., 1999) and dissolved minerals which are not considered a risk to health, but which do affect the palatability of the water (Thompson, 2003). In a 2011 survey of rural Saskatchewan residents 25% of respondents reported having concerns the taste, odor, colour, or cloudiness of their tap water (McLeod et al., 2015).

The hypothesis motivating this study was that exposure to poor water quality drinking water, with respect to health standards and/or aesthetic objectives, could increase the risk of hypertension, ischemic heart disease, or stroke for residents in some areas of rural Saskatchewan. The primary goal of the present study was to use an ecological design to investigate associations between groundwater quality and hypertension, IHD, and stroke in rural Saskatchewan using existing water quality surveillance and population-based administrative health data. The first objective was to examine associations between groundwater arsenic concentration and hypertension, IHD, and stroke in rural Saskatchewan. The final objectives were to examine the associations between both groups of substances monitored as health standards and substances measured as aesthetic objectives and the prevalence of hypertension, IHD and stroke in rural Saskatchewan.

## **6.3 Materials and Methods**

### *6.3.1 Study Area*

The study area was limited to the southern part of Saskatchewan (Figure 6.1); because the population and water quality data were very sparse, the northern part of the province was excluded from the study. The study area extent corresponds to the borders of rural municipalities (RMs), an administrative unit that provided the basis for division of the study area into geographic units for analysis; the east, west and southern borders of the study area also correspond to Saskatchewan provincial borders.



**Figure 6.1** A map showing the province of Saskatchewan and the study area for the analysis of associations between water quality and hypertension, IHD and stroke along with the rural municipalities within the study area prior to aggregation. The small inset map provides context regarding the location of Saskatchewan within Canada.

### *6.3.2 Geographic Units for Analysis*

Rural municipalities provided the basis for division of the study area into geographic units for analysis. In the administrative health data used for the study, residents were assigned a residence code based on place of residence or mailing address. The first 3 digits of the residence code was used to assign study subjects to the RMs. Because the population in some RMs is small and the counts of cases were stratified by sex and age category, RMs with fewer than 500 residents over the age of 19 years were aggregated to minimize the occurrence of zero counts. All RMs were aggregated according their RM number, but because they have been combined over time for administrative purposes, they are not numbered consecutively and there are occasionally large

differences in numbers of adjacent RMs. An RM with an adult population <500 was aggregated with an adjacent lower numbered RM where the difference in RM number was <3; if the RM numbers differed by 3 or more the RM was aggregated with the next higher numbered RM.

First Nations reserves were also aggregated with RMs. Because First Nations often have more than one reserve which can be geographically distant, the most populous reserve for each First Nation according to the 2006 Census of Canada (Statistics Canada, 2014) was identified. Census Consolidated Subdivision (CCS) correspond to RMs and are numbered according the corresponding RM; therefore, the CCS associated with the most populous reserve was used to match each First Nation to an RM.

Cities were excluded from the geographic units used in the analysis. For the purposes of the study, the rural population was considered all those living outside cities. Cities can be incorporated in Saskatchewan once the population of a center reaches 5000.

### *6.3.3 Disease Data*

Because Saskatchewan has universal health care coverage, a database of all residents holding health care coverage is maintained by the Ministry of Health for administrative purposes. Using de-identified data and in a secure facility, cohorts of all Saskatchewan residents age 35-74 covered by Saskatchewan health care as of June 30 were extracted each year from 2004 through 2010.

All hospital and physician visits are billed to the province. International Classification of Diseases (ICD) codes in administrative data were used to identify cases for each year from 2006-2010 based on previously validated algorithms for hypertension (Tu et al., 2007; Quan et al., 2009, Robitaille et al., 2012, Quan et al., 2013, IHD (Tu et al., 2010; Robitaille et al., 2013), and stroke (Moore et al., 2008, Tu et al., 2013) (Table 6.1). Cases of hypertension that could reflect pregnancy-related hypertension were excluded (Table 6.1). Once a study subject was identified as a case, they were considered a prevalent case in all subsequent years.

The algorithms for hypertension and IHD are used by the Canadian Chronic Disease Surveillance System (CCDSS) (Government of Canada 2014). The algorithm for stroke differed slightly from that currently used by the CCDSS and was based on the algorithm suggested by Tu et al. (2013) that included the diagnosis of transient ischemic attack and incorporated physician

visits because reliance only on hospital records was found to underestimate the prevalence of stroke. In addition, to make the algorithm more specific for stroke, ICD-9 code 433 was excluded from the definition (Moore et al., 2008). Stroke cases were counted as prevalent in all subsequent years, which also differs from the CCDSS algorithm which considers stroke prevalence on an annual basis.

**Table 6.1** Case definitions and International Classification of Diseases (ICD) codes used to identify cases of hypertension, ischemic heart disease, and stroke, as well as chronic obstructive pulmonary disease for use as a covariate, in Saskatchewan administrative health data.

Case Definition	ICD codes	Exclusions
<b>Hypertension</b>		
2 physician claims within 2 years Or 1 hospital discharge abstract record	<u>ICD-9<sup>1</sup></u> : 401, 402, 403, 404, 405 <u>ICD-10<sup>2</sup></u> : I10, I11, I12, I13, I15	Pregnancy induced hypertension: exclude any cases with any obstetrical code 120 before or 180 days after: <u>ICD-9<sup>1</sup></u> : 641-676, V27 <u>ICD-10<sup>2</sup></u> : O1, O21-95, O98, O99, Z37
<b>Stroke / Transient ischemic attack</b>		
2 physician claims within 1 year Or 1 hospital discharge abstract	<u>ICD-9<sup>1</sup></u> : 362.3, 430, 431, 434, 436, 435 <u>ICD-10<sup>2</sup></u> : I60, I61, I63, I64, H34.1, G45	none
<b>Ischemic Heart Disease</b>		
2 physician claims within 1 year Or 1 hospital discharge abstract	<u>ICD-9<sup>1</sup></u> : 410, 411, 412, 413, 414 <u>ICD-10<sup>2</sup></u> : I20, I21, I22, I23, I24, I25 <b>Procedure codes</b> <i>Percutaneous coronary intervention</i> <u>CCP</u> : 48.02, 48.03 <u>ICD-9 CM</u> : 36.01, 36.02, 36.05 <u>CCI</u> : 1.IJ.50, 1.IJ.57.GQ, 1.IJ.54 <i>Coronary artery bypass graft</i> <u>CCP</u> : 48.11-48.19 <u>ICD-9 CM</u> : 36.10-36.19 <u>CCI</u> : 1.IJ.76	none
<b>Chronic Obstructive Pulmonary Disease</b>		
≥ 1 physician claims within 1 year Or ≥ 1 hospital discharge abstract	<u>ICD-9<sup>1</sup></u> : 491, 492, 496 <u>ICD-10<sup>2</sup></u> : J41, J42, J43, J44	none

<sup>1</sup>International Classification of Diseases 9th Edition used to identify cases in Physician Services Claims File: Medical Services Branch.

<sup>2</sup>International Classification of Diseases 10<sup>th</sup> Edition used to identify cases in Hospital Discharge Abstract Database.

CCP = Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures, ICD-9 CM = International Classification of Diseases 9th Edition Clinical Modification, CCI = Canadian Classification of Health Interventions



The exposure-outcome analysis considered prevalent cases for each of 5 one-year periods starting in 2006 and ending in 2010. The administrative health databases were examined for cases beginning in 2002 to allow detection of all prevalent cases identified with a minimum 4 year run in period.

Counts of prevalent cases of hypertension, IHD, and stroke were stratified for each geographic unit by sex and age category (34-44, 45-54, 55-64 and 65-74). Ethics approval was obtained from the University of Saskatchewan Behavioral Research Ethics Board (Bio 12-332)

#### *6.3.4 Water Data*

The process for assessing the mean exposure for each water quality measure of interest for each geographical unit in the study area has been described in detail elsewhere (Chapter 4). The methods are briefly summarized here. The Saskatchewan Water Security Agency provided access to water quality surveillance data from both public water supplies and private wells. The public water supply data included repeated samples taken for routine monitoring of public water supplies across Saskatchewan over the period 1985-2012. The private well data were de-identified to protect well owners' privacy; wells were sampled from 1996 to 2011 as part of the Water Security Agency's Rural Water Quality Advisory Program. This innovative program provided a resource to private water supply owners, who could access subsidized water quality testing and water supply management advice on a voluntary basis. The private well data were de-identified and well locations were generalized to the nearest section of land (a parcel approximately 1.6 km x 1.6 km) to protect the confidentiality of program participants.

Water analysis included trace metals and major ions that are listed in the Saskatchewan Drinking Water Quality Standards and Objectives (Water Security Agency n.d.). Only those measured routinely for both public supplies and private were retained for inclusion in the exposure assessment. Trace metals were excluded where reported concentrations were below detection limit for more than 75% of samples in either type of supply. Health standards measures retained for exposure assessment included concentrations of arsenic, barium, boron, lead, nitrate, selenium and uranium. The aesthetic objectives measures retained for exposure assessment included reported alkalinity, chloride, copper, hardness, iron, magnesium, manganese, sodium, sulphate, total dissolved solids, and zinc concentrations.

For samples that were below detection limits, concentrations were imputed as  $\frac{1}{2}$  the detection limit for the analytical method in use at the time of sampling. The data for all parameters were right skewed and log transformed to approximate a normal distribution. Public water supply data were analyzed separately from the private well data.

The public water supply data consisted of repeated measures at each location. However, empirical Bayesian kriging (the geostatistical technique used to interpolate measures between point locations) required a single value for each location. To account for the repeated measures, linear mixed models with random effects for site and a structured error term incorporating time between samples were developed for each water parameter to estimate a single measure for each public water supply site using PROC MIXED (SAS Institute Inc., Cary, NC, USA).

Because the private well locations had been generalized to the centroid of the section of land on which they were located for confidentiality purposes, their locations were separately slightly by alternately adding or subtracting increments of 10 m to their latitude and longitude to create unique locations for each well to facilitate kriging.

Principal components analysis (PCA) (SAS Institute Inc., Cary, NC, USA) was used to summarize groups of standards and objectives for each type of water supply (Chapter 4).

Principal components (PC) with eigenvalues  $>1$  were retained. Varimax rotation was employed to maximize the amount of variability explained by each PC, and the PC coefficients were used to calculate PC scores for each water supply.

Arsenic concentrations and PC scores were interpolated across the study area using Empirical Bayesian kriging and summarized at a resolution of 800m x 800m (ArcGIS, ESRI, Redlands, CA) (Chapter 4). Mean values of the predicted logged arsenic concentration each principal component score were extracted from public supply and private well data over each geographic unit for use as exposure variables in the ecologic analysis (ArcGIS, ESRI, Redlands, CA). The logged arsenic concentrations were back transformed to concentration in  $\mu\text{g/L}$  for the exposure-outcome analysis.

### 6.3.5 Covariates

The prevalence of chronic obstructive pulmonary disease (COPD) was used as a covariate in the models as a proxy for smoking rates. COPD cases were identified from administrative health data in each year starting in 2002 to allow a minimum 4 year run in period to identify prevalent cases from 2006 to 2010 using ICD codes according to a published algorithm (Gershon et al., 2009) (Table 6.1). COPD prevalence, stratified by sex and age category was calculated for each geographic unit for 2006-2010.

Covariate data for education and total income were accessed from publically available reporting of the 2006 Census of Canada (Statistics Canada, 2014). Census data were available at the level of census subdivision, and were aggregated to obtain covariate data corresponding to the geographic units used in the analysis. The proportion of residents not attaining the level of high school certificate was calculated and stratified by sex and age category. However, the educational attainment data was subject to random rounding and suppression for areas with a population < 40 (Statistics Canada, 2008a). After aggregation there were areas where the proportion not completing high school was > 1 due to rounding. Due to suppression there were some areas for which the population at risk was zero. For strata where the proportion not completing high school was 0 or >1, the average value for that sex and age category for all residents of Saskatchewan not living in cities was imputed.

Average total income was calculated and aggregated for each geographic unit from the census subdivision data. Income data were only available for wider age intervals than those used for the disease data. Therefore, the average total income for those aged 25-44 was assigned to the 35-44 age category, income for those aged 45-64 was assigned to the 45-54 and 55-64 age categories, and income for those 65 and over was used for the 65-74 age category. The income data were subject to suppression for populations less than 250 (Statistics Canada, 2008b). Where aggregated income data was missing for a sex and age category the average total income of the corresponding age and sex category for all Saskatchewan residents not living in cities was imputed. The average total income was centered on the mean value for Saskatchewan residents in the corresponding age and sex category not living in cities, and scaled so that the regression coefficients for average income were interpretable as change in disease prevalence relative to a change in thousands of dollars of personal income.

### 6.3.6 Bayesian Statistical Analyses of Associations between Water Quality Data and Outcomes

Associations between exposure, measured as arsenic concentrations, health standards PC scores, and aesthetic objective PC scores, and the outcomes of interest, including counts of prevalent cases of hypertension, IHD and stroke, were investigated using generalized linear mixed models (GLMMs) with a log link function. Counts of cases for each outcome were assumed to follow a Poisson distribution with mean  $\mu_i = E_i \theta_i$  where  $E_i$  is the expected count, and  $\theta_i$  is the standardized morbidity ratio (SMR).

Expected case counts stratified by sex and age categories, were calculated for hypertension, IHD, and stroke prevalence for each year from 2006-2010. The study area-wide prevalence for each of the outcomes was calculated separately for each age stratum for women and for men. The total number of cases across all geographic units for each stratum was divided by the total population at risk for each stratum. The stratum-specific expected number of prevalent cases was then calculated for each geographic unit by multiplying the population at risk in each stratum for each area by the study wide prevalence for that stratum.

The 2010 prevalence data for each outcome were analyzed using Bayesian models based on the hierarchical model developed by Besag, York and Mollie (1991) (Equation 6.1). The models incorporated a spatially correlated random effect as well as an unstructured independent random effect. An intrinsic conditionally autoregressive prior distribution was used to model the spatially structured random effect ( $v_i$ ). Each spatial random effect was assumed to follow a normal distribution whose mean and precision are conditioned on the random effects of neighboring areas, where  $\delta_i$  is the set of adjacent neighbors for the  $i$ th geographic unit and  $n_{\delta_i}$  is the number of neighbors for area  $i$ . A zero-mean Gaussian distribution was used to model the unstructured random effects ( $h_i$ ). The precision on the structured ( $\tau_v$ ) and unstructured ( $\tau_h$ ) random effects were given gamma (0.001, 0.001) hyperpriors. A zero mean Gaussian distribution was assumed for the regression coefficients ( $\beta_{1...k}$ ) for each of  $k$  fixed effects ( $x_{1...k}$ ) and for the intercept ( $\beta_0$ ), and these were assigned uninformative priors. Because estimates are conditioned on information from neighboring areas, this model allows for smoothing of estimated counts, and is particularly useful for area data where small counts can lead to spurious extreme values for calculated relative risks (Lawson, 2013).

$$\log \mu_i = \log E_i + \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + h_i + v_i \quad (6.1)$$

$$\beta_0, \beta_{1\dots k} \sim \text{Normal}(0, 10^{-5})$$

$$h_i \sim \text{Normal}(0, \tau_h)$$

$$v_i | \mathbf{v}_{-i} \sim \text{Normal}(\bar{v}_{\delta_i}, \tau_v/n_{\delta_i})$$

$$\bar{v}_{\delta_i} = \frac{1}{n_{\delta_i}} \sum_{j \in \delta_i} v_j$$

$$\tau_h, \tau_v \sim \text{gamma}(0.001, 0.001)$$

OpenBUGS 3.2.3 (Lunn et al., 2009) was used for the Bayesian modeling and also to generate the adjacency matrix specifying the neighborhood structure based on queen contiguity for the spatial random effects. Separate models were built to investigate each of the associations between the exposure measures, including arsenic concentrations, health standards PC scores, and aesthetic objectives PC scores for public or private water supplies, and the outcomes of interest, including hypertension, IHD and stroke, for a total of six models for each outcome. The covariates included in each model were stratum-specific COPD prevalence, proportion of residents not completing high school, and average total income in thousands of dollars. The water quality risk factors were initially evaluated as continuous variables. If the 95% credible interval for a water-related variable excluded 1 in the Bayesian analysis for 2010, the linearity assumption was evaluated by adding a quadratic term for that variable to the model to assess if the credible interval for the squared term included 1. Furthermore, a third model was assessed with that variable categorized into quintiles to further characterize the potential association between the exposure variable and outcome.

For each Bayesian model, three chains were initiated and the Brooks-Gelman-Rubin diagnostic (Gelman and Rubin, 1992) was used to evaluate convergence. Convergence was assessed both visually using plots and quantitatively using the CODA (Plummer et al., 2006) package in R 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria); convergence was considered satisfactory once the 97.5% upper bound of the scale reduction factor was 1.05 or less for each parameter. Once convergence was reached, 20,000 iterations from each of three chains were

sampled for a total of 60,000 iterations to obtain estimates for the model parameters. Adequate sampling for each parameter was confirmed by ensuring that the MCMC was less than 5% of the sample standard deviation.

#### *6.3.7 Frequentist Statistical Analyses of Associations between Water Quality Data and Outcomes*

Frequentist GLMMs using a log link and a single independent random effect for geographic unit in PROC GLIMMIX (SAS Institute Inc., Cary, NC, USA) was also used to analyze data from 2006-2010 for each outcome. Results were compared for the yearly frequentist models between the 2010 frequentist and Bayesian models to assess consistency in the modeled relationships. Where no associations were identified in the 2010 Bayesian model, the results of frequentist models were identified for further examination and discussion if a particular exposure association was significant over at least 4 of the 5 examined years.

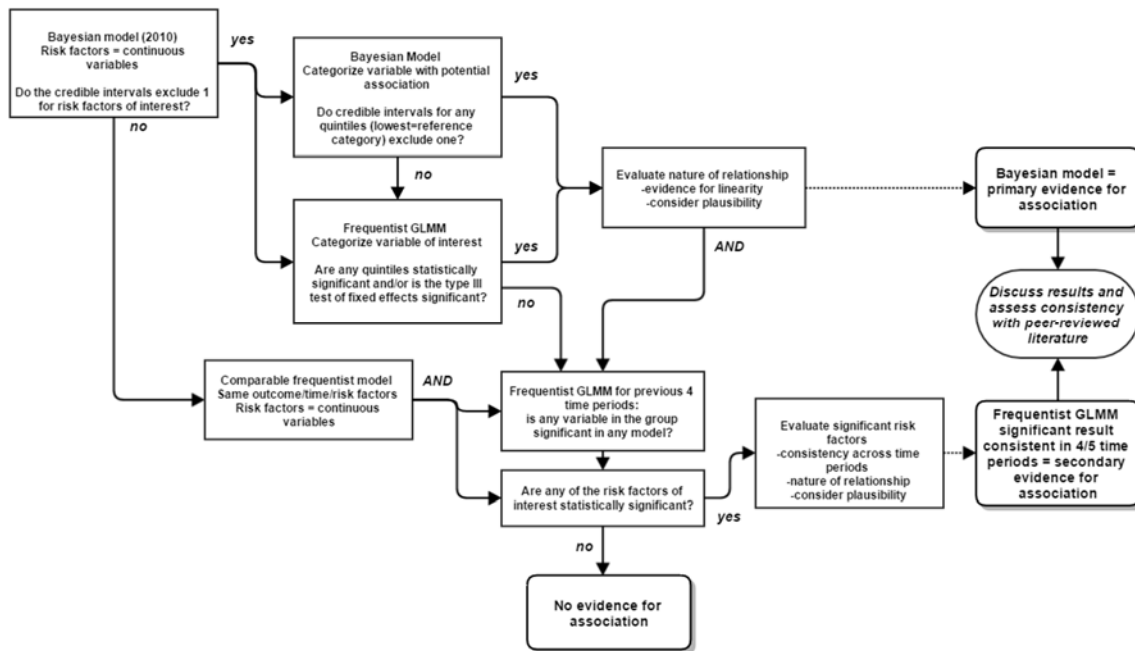
#### *6.3.8 Model Results and Assessment of Fit*

From each of the models, regression coefficients were exponentiated for interpretation as risk ratios (RR), which represented a multiplicative effect of each of the risk factors on the SMR. Point estimates for the RR along with 95% credible intervals (CrI) for Bayesian models and 95% confidence intervals (CI) for frequentist models were reported.

Because the frequentist models did not account for spatial relationships, the presence of spatial correlation in the random effects residuals for the 2010 frequentist models was examined by calculating Moran's I in GeoDA 1.8.10 (The Center for Spatial Analysis, Chicago, Illinois). First order queen's contiguity was specified for the form of the spatial structure for the Moran's I calculation.

Where potential associations were identified in the Bayesian models, the deviance Information criterion (DIC), a measure of model fit, was examined to assess whether model fit was improved by the inclusion of spatial random effects. The DIC was compared between the Bayesian model including both unstructured and spatially structured random effects and a model with the spatial random effect excluded. The value of pD was also examined to ensure the DIC comparison was valid. The pD is a measure of model complexity representing the effective number of parameters in a model (Spiegelhalter et al., 2002). If pD takes a negative value, the DIC is not interpretable.

Because multiple models were used to evaluate associations between the drinking water risk factors and the outcomes, a decision tree approach was used to guide examination of the results from the Bayesian and frequentist models for 2010 and the frequentist models for 2006-2009 to identify results for further discussion (Figure 6.2). For public water supplies, drinking water arsenic concentrations, three health standard PC scores, and four aesthetic objectives PCs were evaluated as risk factors. For private wells, arsenic concentrations, three health standard PC scores, and three aesthetic objectives PCs were evaluated. Consequently, a total of 15 risk factors were assessed for associations with each of hypertension, IHD, and stroke prevalence. A conservative experiment-wise error rate for the Bayesian modelling would be based on 15 tests for each outcome. The probability of falsely identifying an observed number of associations given 15 tests was calculated using publically available software (FreeCalc, EpiTools, AusVet Animal Health Services) specifying a modified hypergeometric distribution. As an example, the probability of falsely identifying a single association in the analysis can be estimated as an experiment-wise error =  $1-(1-\alpha)^k$  for  $k$  tests. For situations where significant associations were identified in the Bayesian analysis, the software estimated the probability of finding at least the observed number of associations assuming no true association existed. This calculation was used to provide context to the discussion on the potential for experiment-wise error. Finally, results that were consistent across 2 or more of hypertension, IHD and stroke were considered to have more substance than results that applied to only one of the three outcomes.



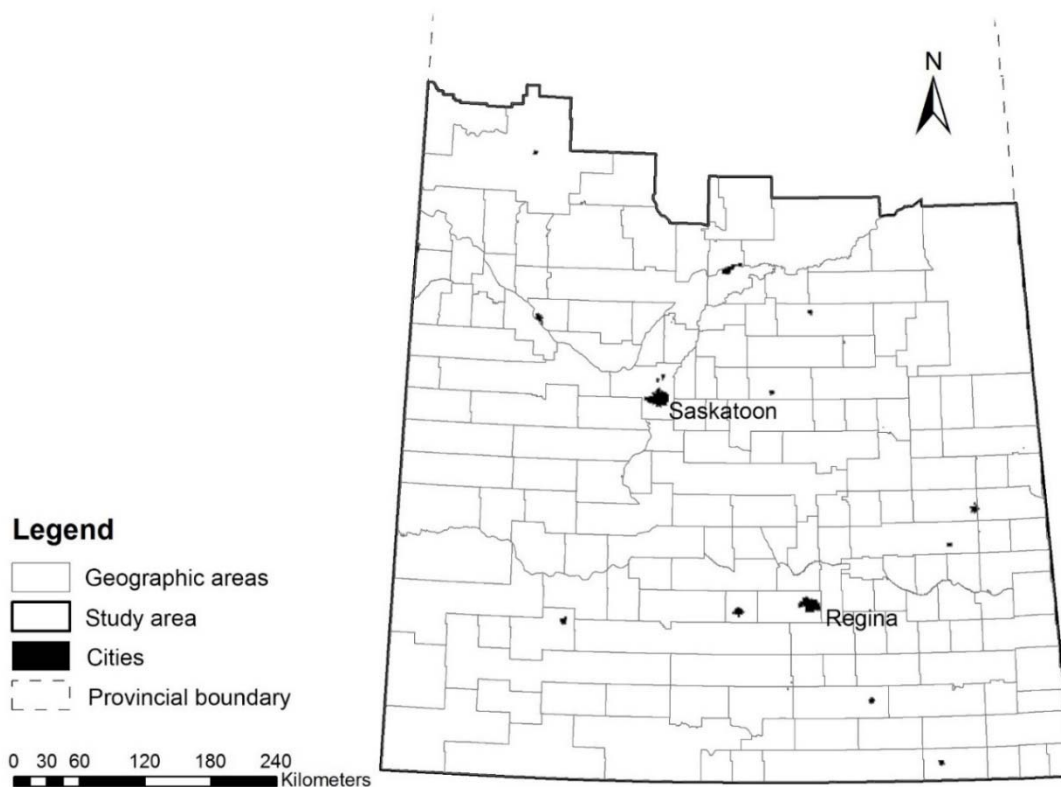
**Figure 6.2** A flow chart outlining process of evaluating potential associations between water quality risk factors and hypertension, ischemic heart disease and stroke in Bayesian and frequentist models.

Sensitivity to hyperprior choice for the random effects was evaluated by comparing the effect of different hyperprior specifications on the association between the health standards PC scores from public water supplies and stroke prevalence for 2010. The effect estimates for the PC<sub>scores</sub> were compared between the reported model with a gamma (0.001, 0.001) distribution for the precision on each of the random effects, a model with a gamma (0.5, 0.0005) distribution placed on the precision for each of the random effects, and a uniform (0, 5) distribution placed on the standard deviation for the precision, where precision is equal to the inverse of the squared standard deviation (Jang et al., 2007; Lawson, 2013)

## 6.4 Results

The geographic units used in the analysis were comprised of 168 areas resulting from aggregation of the 296 RMs present within the study area (Figure 6.3). Stratification of 168 areas by sex and four age categories resulted in 1344 strata for use in each of the models.





**Figure 6.3** A map of Saskatchewan showing geographic regions formed by amalgamation of RMs for analysis of the association between measures of water quality and hypertension, ischemic heart disease, and stroke.

Within the study area, the population of residents aged 35-74 for whom location information was available ranged from 178,324 to 188,423 for 2006 to 2010 respectively. The population increased between 2006 and 2010 for the age groups from 45-74, but decreased for the 35-44 age group. In 2010, the proportion residents in the 35-44 age group was 23.4% (44,163/188,423), while the proportion was 31.7% (59,718/188,423) in the 45-54 age group, 27.6% (51,932/188,423) in the 55-64 age group, and 17.3% (32,610/188,423) in the 65-74 age group. The percent of women in the population was 48.7% (91,742/188,423) in 2010. The 2010 population at risk within geographic units ranged from 316 to 6390 (median=913).

The observed prevalence of hypertension, IHD and stroke in residents aged 35-74 within the study area increased over the course of the study period (Table 6.2). The prevalence of hypertension for all residents aged 35-74 among individual geographic units for 2010 ranged

from a minimum of 20.1% to a maximum of 43.4%, with a median prevalence of 29.8%. The observed prevalence was similar among males (29.4%) and females (29.8%), and prevalence ranged from 8.8% in 35-44 year olds to 59.6% among 65-74 year olds.

The observed prevalence of IHD in 2010 among geographic units ranged from 2.4% to 14% with a median prevalence of 6.4%. The observed prevalence in 2010 was 8.3% for males and 4.6% for females, and ranged from 0.9% in 35-44 year olds to 17.2% among 65-74 year olds.

The observed prevalence of stroke among geographic units ranged from a minimum 1.4% to a maximum of 5.2%, with a median prevalence of 3.2%. Observed prevalence was similar among males (3.4%) and females (3.1%) and ranged from 0.9% in 35-44 year olds to 8.6% among 65-74 year olds.

**Table 6.2** Hypertension, ischemic heart disease, and stroke prevalence in the study area population ages 35 to 75 for those not living in a Saskatchewan city.

Year	Study population (n)	Hypertension		IHD		Stroke	
		cases (n)	Prevalence (%)	Cases (n)	Prevalence (%)	Cases (n)	Prevalence (%)
2006	178 324	45285	25.4	9117	5.1	4220	2.4
2007	179 496	48187	26.8	10100	5.6	4824	2.7
2008	183 250	50831	27.7	10788	5.9	5342	2.9
2009	185 154	53406	28.8	11609	6.3	5789	3.1
2010	188 423	55791	29.6	12240	6.5	6141	3.3

In the raw public water supply data, the median and 95<sup>th</sup> percentile of observed arsenic concentrations were 0.9 µg/L and 14 µg/L; after summarizing values for each site using GLMM the median and 95<sup>th</sup> percentile for predicted mean values across sites were 0.8 µg/L and 7.5 µg/L. The median and 95<sup>th</sup> percentile for observed arsenic concentrations from private wells were 0.9 µg/L and 23 µg/L. Fewer samples exceeded the Saskatchewan drinking water standard of 10 µg/L from public water supplies (6.9 %) than from private wells (13.5 %). Among the public water supply samples, 22.9 % were below detection limits, while 21.3 % were below detection limits among the private well samples (Chapter 4).

After kriging, predicted arsenic concentrations were averaged within each of the 168 aggregated RMs and the median arsenic exposure from public water supplies was 0.9 µg/L (95<sup>th</sup> percentile = 1.9 µg/L), while for private wells the median arsenic exposure was 1.3 µg/L (95<sup>th</sup> percentile = 3.1

µg/L). The cut points defining the quintiles of arsenic concentration for the public water supply data were 0.73 µg/L, 0.88 µg/L, 1.01 µg/L, and 1.31 µg/L. The cut points for the quintiles of arsenic concentrations from private wells were 0.86 µg/L, 1.16 µg/L, 1.55 µg/L, and 2.25 µg/L.

PCA for each of the public and private water supply data sets yielded three health standards principal components ( $PC_{\text{health}}$ ) with eigenvalues greater than one (Table 6.3). PCA for the aesthetic objectives ( $PC_{\text{aesthetic}}$ ) resulted in different numbers of components for the public water supplies and private wells, with four PCs retained for public supplies and three for private wells (Table 6.3). The detailed results of this analysis were reported in Chapter 4.

**Table 6.3** Summary of principal components analysis of groundwater parameters for health standards and aesthetic objectives from public water supplies and private wells. The loadings for each component are shown along with the eigenvalue and cumulative variance explained for each retained component; the maximum loading for each parameter is indicated in bold. Details of the PCA analysis can be found in Chapter 4.

	Public Water Supplies				Private Wells		
<b>Health Standards</b>							
	PC1 <sub>health</sub>	PC2 <sub>health</sub>	PC3 <sub>health</sub>		PC <sub>health</sub>	PC2 <sub>health</sub>	PC3 <sub>health</sub>
Arsenic	-0.121	0.142	<b>0.808</b>		-0.341	-0.091	<b>0.474</b>
Barium	0.047	<b>-0.818</b>	-0.141		-0.041	<b>0.893</b>	0.100
Boron	-0.062	<b>0.903</b>	-0.123		-0.195	<b>-0.818</b>	0.168
Lead	<b>0.472</b>	0.092	0.171		0.156	0.026	<b>0.893</b>
Nitrate	<b>0.768</b>	-0.071	-0.164		<b>0.770</b>	0.275	-0.110
Selenium	<b>0.867</b>	-0.220	0.019		<b>0.853</b>	-0.007	0.074
Uranium	0.387	-0.290	<b>0.576</b>		<b>0.772</b>	-0.013	-0.049
Eigenvalue	2.127	1.275	1.059		2.290	1.381	1.057
Cumulative variance (%)	30.4	48.6	63.7		32.7	52.5	67.6
<b>Aesthetic Objectives</b>							
	PC1 <sub>aesthetic</sub>	PC2 <sub>aesthetic</sub>	PC3 <sub>aesthetic</sub>	PC4 <sub>aesthetic</sub>	PC1 <sub>aesthetic</sub>	PC2 <sub>aesthetic</sub>	PC3 <sub>aesthetic</sub>
Alkalinity	<b>0.755</b>	0.111	0.164	-0.198	<b>0.687</b>	-0.023	0.217
Chloride	<b>0.753</b>	-0.193	0.002	0.226	<b>0.779</b>	0.043	-0.195
Copper	0.127	0.012	-0.200	<b>0.714</b>	0.030	0.223	<b>-0.757</b>
Hardness	0.009	<b>0.973</b>	0.066	0.042	0.067	<b>0.960</b>	0.038
Iron	0.138	-0.089	<b>0.901</b>	0.053	0.117	0.121	<b>0.784</b>
Magnesium	-0.014	<b>0.961</b>	0.055	0.038	0.103	<b>0.951</b>	0.020
Manganese	0.188	0.452	<b>0.711</b>	-0.065	0.062	0.468	<b>0.663</b>
Sodium	<b>0.914</b>	-0.199	0.136	0.026	<b>0.922</b>	-0.116	0.118
Sulphate	<b>0.663</b>	0.517	0.018	0.116	<b>0.609</b>	0.555	0.076
Total dissolved solids	<b>0.920</b>	0.288	0.121	-0.016	<b>0.907</b>	0.325	0.082
Zinc	-0.089	0.078	0.237	<b>0.763</b>	-0.091	<b>0.396</b>	-0.375
Eigenvalue	3.746	2.362	1.264	1.181	3.775	2.184	1.779
Cumulative variance (%)	34.1	55.5	67.0	77.8	34.3	54.2	70.4

PC scores were calculated for each water supply and summarized (Table 6.4). The scores were subject to kriging and the predicted scores averaged for each geographic unit to derive the measure of exposure for the health standard and aesthetic objective PC scores. The resulting

averaged predicted scores used to measure exposure over each geographic unit of analysis had a smaller range of PC scores than that observed for the scores for each site used as input for kriging (Table 6.4)

**Table 6.4** Median, 5<sup>th</sup> and 95<sup>th</sup> percentile values for the mean exposures for principal component scores estimated for each geographic unit (left) and the scores for individual water supplies used as input for kriging for public water supplies and private wells.

Mean area exposure measures (after kriging)					PCA results (point values before kriging)			
<i>Public water supplies</i>								
Variable	n	P5	Median	P95	n	P5	Median	P95
PC1 <sub>health</sub>	168	-0.55	0.09	0.52	459	-1.47	0.01	1.83
PC2 <sub>health</sub>	168	-0.66	0.04	0.69	459	-1.68	0.18	1.49
PC3 <sub>health</sub>	168	-0.62	0.05	0.68	459	-1.60	0.08	1.59
PC1 <sub>aesthetic</sub>	168	-0.63	0.02	0.68	435	-1.66	0.11	1.59
PC2 <sub>aesthetic</sub>	168	-0.78	0.07	0.65	435	-2.32	0.23	1.13
PC3 <sub>aesthetic</sub>	168	-0.56	-0.07	0.63	435	-1.74	0.08	1.55
PC4 <sub>aesthetic</sub>	168	-0.48	0.01	0.41	435	-1.56	-0.06	1.81
<i>Private wells</i>								
PC1 <sub>health</sub>	168	-0.55	0.05	0.63	3970	-1.51	-0.09	1.73
PC2 <sub>health</sub>	168	-0.70	-0.15	0.73	3970	-1.48	-0.09	1.70
PC3 <sub>health</sub>	168	-0.30	0.17	0.57	3970	-1.67	0.09	1.39
PC1 <sub>aesthetic</sub>	168	-0.69	0.12	0.78	3999	-1.72	0.06	1.53
PC2 <sub>aesthetic</sub>	168	-0.91	0.18	0.61	3999	-2.20	0.15	1.23
PC3 <sub>aesthetic</sub>	168	-0.41	0.07	0.50	3999	-1.73	0.09	1.49

P5 = 5<sup>th</sup> percentile, P96 = 95<sup>th</sup> percentile

Among geographic units, the observed COPD prevalence in 2010 for all residents aged 35-74 ranged from a minimum of 2.7% to a maximum of 10.9%, with a median of 5.5%. From 2006-2010 the observed prevalence was lowest in 2006 with a median of 3.7% among geographic units, increasing each year through 2010.

Based on Census of Canada 2006 results, of the population age 35-74 in the study area not living in cities, the median proportion not completing high school was 29.6% (5<sup>th</sup> percentile = 17.0% and 95<sup>th</sup> percentile = 42.9%). The median average total personal income reported for residents in the study area not living in cities was \$27,375 (5<sup>th</sup> percentile = \$19,893 and 95<sup>th</sup> percentile = \$37,169).

#### *6.4.1 Hypertension and arsenic concentrations*

No associations between groundwater arsenic concentration and hypertension prevalence were evident in the Bayesian model for 2010 for public water supplies (Table 6.5). In the frequentist model for the 2010 data from public water supplies, an increase in arsenic concentration was associated with a decrease in hypertension prevalence (RR=0.966, 95% CI 0.935-0.999). For the private well data, there was no association between arsenic concentration and hypertension prevalence in the Bayesian or the frequentist model for 2010 (Table 6.5).

**Table 6.5** Comparison of Bayesian and frequentist models for the association between arsenic concentrations in drinking water and prevalence of hypertension for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.034	0.987	1.083	Intercept	1.037	0.994	1.082	0.09
Arsenic (µg/L)	0.972	0.936	1.011	Arsenic (µg/L)	<b>0.966</b>	<b>0.935</b>	<b>0.999</b>	<b>0.04</b>
COPD prevalence	1.042	0.845	1.288	COPD prevalence	1.046	0.847	1.292	0.68
Education <sup>1</sup>	0.976	0.923	1.032	Education <sup>1</sup>	0.983	0.929	1.040	0.55
Income <sup>2</sup>	1.000	0.999	1.002	Income <sup>2</sup>	1.000	0.999	1.001	0.77
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.055	0.013	Unstructured RE		0.008	0.001	
Spatially structured RE		0.070	0.012					
Burn in period = 80,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.025	0.984	1.067	Intercept	1.012	0.973	1.053	0.55
Arsenic (µg/L)	0.987	0.966	1.009	Arsenic (µg/L)	0.993	0.973	1.013	0.47
COPD prevalence	1.041	0.844	1.284	COPD prevalence	1.043	0.844	1.288	0.70
Education <sup>1</sup>	0.976	0.922	1.032	Education <sup>1</sup>	0.984	0.930	1.041	0.57
Income <sup>2</sup>	1.000	0.999	1.002	Income <sup>2</sup>	1.000	0.999	1.001	0.77
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.054	0.012	Unstructured RE		0.008	0.001	
Spatially structured RE		0.073	0.011					
Burn in period = 80,000 iterations								
Number of observations = 1344								

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

In the frequentist GLMMs for 2006-2009, increased arsenic concentration in public water supplies was associated with a decreased prevalence of hypertension in 2006 (RR=0.949, 95% CI 0.912-0.987), 2007 (RR=0.960, 95% CI 0.926-0.997), and 2008 (RR=0.965, 95% CI 0.932-1.000) but not in 2009 (Table 6.6). When arsenic concentrations from public supplies were categorized into quintiles, the relationship between the quintiles of arsenic concentrations and hypertension was clearly not monotonic with a similar pattern demonstrated in all years (results not shown). No associations between groundwater arsenic concentration in private wells and hypertension were demonstrated in the frequentist models for 2006-2009 (Table 6.6).

**Table 6.6** Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and prevalence of hypertension for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	<b>2009</b>				<b>2008</b>				<b>2007</b>				<b>2006</b>			
	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	1.036	0.991	1.083	0.12	1.035	0.990	1.082	0.14	1.039	0.991	1.089	0.11	1.050	0.999	1.103	0.06
Arsenic (µg/L)	0.968	0.935	1.002	0.07	<b>0.965</b>	<b>0.932</b>	<b>1.000</b>	<b>0.048</b>	<b>0.960</b>	<b>0.926</b>	<b>0.997</b>	<b>0.03</b>	<b>0.949</b>	<b>0.912</b>	<b>0.987</b>	<b>0.01</b>
COPD prevalence	0.977	0.783	1.219	0.84	0.989	0.778	1.256	0.93	0.931	0.721	1.202	0.58	0.866	0.651	1.152	0.32
Education <sup>1</sup>	0.996	0.940	1.055	0.88	1.006	0.949	1.067	0.84	1.018	0.959	1.081	0.56	1.031	0.969	1.097	0.33
Income <sup>2</sup>	1.000	0.999	1.001	0.76	1.000	0.999	1.001	0.90	1.000	0.999	1.001	0.91	1.000	0.998	1.001	0.84
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.009		0.001		0.009		0.001		0.010		0.002		0.012		0.002	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	1.010	0.969	1.052	0.64	1.014	0.972	1.057	0.52	1.007	0.964	1.052	0.75	1.006	0.960	1.054	0.80
Arsenic (µg/L)	0.994	0.974	1.016	0.60	0.989	0.968	1.011	0.32	0.992	0.970	1.015	0.51	0.992	0.968	1.016	0.50
COPD prevalence	0.975	0.781	1.216	0.82	0.986	0.776	1.252	0.91	0.929	0.719	1.200	0.57	0.865	0.651	1.151	0.32
Education <sup>1</sup>	0.996	0.940	1.056	0.90	1.007	0.950	1.068	0.82	1.019	0.960	1.082	0.54	1.032	0.970	1.098	0.32
Income <sup>2</sup>	1.000	0.999	1.001	0.75	1.000	0.999	1.001	0.89	1.000	0.999	1.001	0.90	1.000	0.998	1.001	0.82
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.009		0.001		0.009		0.001		0.011		0.002		0.012		0.002	

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Total average income in thousands of dollars



#### *6.4.2 Hypertension and health standards PC scores*

In the Bayesian models examining associations between health standards PCs and hypertension prevalence for 2010, no associations were apparent for public water supplies or private wells, with the 95% credible intervals spanning 1 for each of the PC<sub>health</sub> scores (Table 6.7). In the frequentist model for 2010 for public water supplies, an increase in the PC<sub>2health</sub> score was associated with decreased hypertension prevalence (RR=0.935, 96% CI 0.897-0.976). Similarly, an increase in the PC<sub>3health</sub> score was also associated with a decrease in hypertension prevalence (RR=0.951, 96% CI 0.912-0.992). No associations between health standards PCs and hypertension for 2010 were evident for private wells (Table 6.7).

**Table 6.7** Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and hypertension prevalence for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.004	0.982	1.027	Intercept	1.004	0.980	1.028	0.77
PC1 <sub>health</sub>	1.049	0.984	1.121	PC1 <sub>health</sub>	1.035	0.984	1.089	0.18
PC2 <sub>health</sub>	0.965	0.914	1.021	PC2 <sub>health</sub>	<b>0.935</b>	<b>0.897</b>	<b>0.976</b>	<b>0.002</b>
PC3 <sub>health</sub>	0.960	0.914	1.011	PC3 <sub>health</sub>	<b>0.951</b>	<b>0.912</b>	<b>0.992</b>	<b>0.02</b>
COPD prevalence	1.038	0.840	1.279	COPD prevalence	1.048	0.849	1.294	0.66
Education <sup>1</sup>	0.976	0.923	1.033	Education <sup>1</sup>	0.980	0.926	1.037	0.48
Income <sup>2</sup>	1.000	0.999	1.002	Income <sup>2</sup>	1.000	0.999	1.001	0.73
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.055	0.014	Unstructured RE		0.007	0.001	
Spatially structured RE		0.068	0.013					
Burn in period = 100,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.004	0.980	1.028	Intercept	1.001	0.975	1.028	0.93
PC1 <sub>health</sub>	0.993	0.941	1.047	PC1 <sub>health</sub>	0.980	0.934	1.028	0.40
PC2 <sub>health</sub>	0.971	0.928	1.016	PC2 <sub>health</sub>	0.991	0.954	1.029	0.63
PC3 <sub>health</sub>	0.988	0.923	1.057	PC3 <sub>health</sub>	0.992	0.931	1.057	0.81
COPD prevalence	1.042	0.845	1.286	COPD prevalence	1.043	0.844	1.289	0.70
Education <sup>1</sup>	0.976	0.923	1.033	Education <sup>1</sup>	0.984	0.930	1.041	0.57
Income <sup>2</sup>	1.000	0.999	1.002	Income <sup>2</sup>	1.000	0.999	1.001	0.78
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.054	0.013	Unstructured RE		0.008	0.001	
Spatially structured RE		0.075	0.012					
Burn in period = 80,000 iterations								
Number of observations = 1344								

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

In the frequentist GLMM results from 2006-2009 for public water supplies, an increase in the PC1<sub>health</sub> score was associated with an increase in hypertension prevalence in 2006 (RR=1.081, 95% CI 1.017-1.1481) and 2007 (RR=1.066, 95% CI 1.007-1.128), but not in 2008 or 2009 (Table 6.8). An increase in the PC2<sub>health</sub> score was significantly associated with a decrease in hypertension prevalence in 2006 (RR=0.909, 95% CI 0.864-0.955), 2007 (RR=0.912, 95% CI

0.870-0.955), 2008 (RR=0.919, 95% CI 0.879-0.961) and 2009 (RR=0.921, 95% CI 0.881-0.962). PC<sub>2health</sub> was categorized into quintiles and in all years the relationship between the quintiles of PC<sub>2</sub> health was monotonic (results not shown). An association between increased PC<sub>3health</sub> score and decreased hypertension prevalence was also apparent in 2006 (RR =0.942, 95% CI 0.897-0.990) and 2009 (RR=0.954, 95% CI 0.914-0.997), but not in 2007 or 2008 (Table 6.8).

The frequentist GLMM results from 2006-2009 for private wells did not identify any significant associations between any of the PC<sub>health</sub> scores and hypertension prevalence (Table 6.8).

**Table 6.8** Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and hypertension prevalence for 2006-2009 in rural southern Saskatchewan.

	2009				2008				2007				2006				
<i>Public water supplies</i>	95% CI				95% CI				95% CI				95% CI				
	Effect	RR	lower	upper	p	RR	lower	upper	P	RR	lower	upper	p	RR	lower	upper	p
Intercept	1.004	0.979	1.029	0.76	1.000	0.975	1.025	0.97		0.997	0.972	1.024	0.84	0.995	0.968	1.022	0.71
PC1 <sub>health</sub>	1.047	0.993	1.103	0.09	1.047	0.993	1.105	0.09		<b>1.066</b>	<b>1.007</b>	<b>1.128</b>	<b>0.03</b>	<b>1.081</b>	<b>1.017</b>	<b>1.148</b>	<b>0.01</b>
PC2 <sub>health</sub>	<b>0.921</b>	<b>0.881</b>	<b>0.962</b>	<b>&lt;0.001</b>	<b>0.919</b>	<b>0.879</b>	<b>0.961</b>	<b>&lt;0.001</b>		<b>0.912</b>	<b>0.870</b>	<b>0.955</b>	<b>&lt;0.001</b>	<b>0.909</b>	<b>0.864</b>	<b>0.955</b>	<b>&lt;0.001</b>
PC3 <sub>health</sub>	<b>0.954</b>	<b>0.914</b>	<b>0.997</b>	<b>0.03</b>	0.962	0.920	1.005	0.08		0.956	0.913	1.002	0.06	<b>0.942</b>	<b>0.897</b>	<b>0.990</b>	<b>0.02</b>
COPD prevalence	0.982	0.788	1.225	0.87	0.990	0.780	1.256	0.93		0.936	0.726	1.208	0.61	0.877	0.660	1.165	0.36
Education <sup>1</sup>	0.992	0.937	1.051	0.79	1.003	0.946	1.063	0.92		1.015	0.956	1.077	0.63	1.027	0.966	1.093	0.39
Income <sup>2</sup>	1.000	0.999	1.001	0.80	1.000	0.999	1.001	0.97		1.000	0.999	1.001	0.97	1.000	0.999	1.001	0.89
Random effects	Variance		SE		Variance		SE			Variance		SE		Variance		SE	
	0.008		0.001		0.008		0.001			0.009		0.001		0.011		0.002	
<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI				
	Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	1.000	0.974	1.028	0.97	0.996	0.969	1.024	0.78		0.994	0.966	1.023	0.68	0.992	0.962	1.022	0.58
PC1 <sub>health</sub>	0.988	0.940	1.039	0.65	1.003	0.953	1.054	0.92		1.002	0.950	1.057	0.95	1.005	0.949	1.064	0.86
PC2 <sub>health</sub>	0.996	0.957	1.036	0.82	0.997	0.958	1.037	0.87		0.991	0.950	1.034	0.68	0.984	0.941	1.030	0.49
PC3 <sub>health</sub>	1.004	0.939	1.074	0.90	1.004	0.938	1.074	0.92		1.004	0.935	1.079	0.91	1.000	0.926	1.079	0.99
COPD prevalence	0.974	0.780	1.216	0.82	0.985	0.776	1.252	0.90		0.929	0.719	1.199	0.57	0.864	0.650	1.150	0.32
Education <sup>1</sup>	0.996	0.940	1.056	0.90	1.007	0.950	1.068	0.81		1.020	0.960	1.083	0.53	1.033	0.971	1.100	0.30
Income <sup>2</sup>	1.000	0.999	1.001	0.74	1.000	0.999	1.001	0.89		1.000	0.999	1.001	0.91	1.000	0.998	1.001	0.84
Random effects	Variance		SE		Variance		SE			Variance		SE		Variance		SE	
	0.009		0.001		0.009		0.001			0.011		0.002		0.012		0.002	

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

#### *6.4.3 Hypertension and aesthetic objective PC scores*

For public water supplies, no associations were identified between aesthetic objective PC scores and hypertension prevalence in 2010 in either the Bayesian or the frequentist model (Table 6.9). In the Bayesian model for private well aesthetic objective PC scores, an increase in the PC2<sub>aesthetic</sub> score was associated with a decrease in hypertension prevalence in 2010 (RR= 0.941, 95% CrI 0.902-0.982), with a posterior probability that RR<1 of 99.7%. PC2<sub>aesthetic</sub> scores were subsequently categorized into quintiles and no differences in hypertension prevalence were evident between the first and higher quintiles of PC2<sub>aesthetic</sub> where the credible interval did not include 1 (Table 6.9). Similarly, the frequentist model for 2010 suggested that as PC2<sub>aesthetic</sub> scores increased, hypertension prevalence decreased (RR= 0.965, 95% CI 0.932-0.999, p=0.046). When PC2<sub>aesthetic</sub> was categorized, again no significant differences in hypertension prevalence were evident between the first and higher quintiles of PC2<sub>aesthetic</sub>, and the overall type III test of fixed effects for PC2<sub>aesthetic</sub> was not significant (p=0.75).

**Table 6.9** Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and hypertension prevalence for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.003	0.982	1.025	Intercept	0.999	0.976	1.024	0.97
PC1 <sub>aesthetic</sub>	1.041	0.984	1.101	PC1 <sub>aesthetic</sub>	1.013	0.967	1.061	0.59
PC2 <sub>aesthetic</sub>	0.984	0.934	1.034	PC2 <sub>aesthetic</sub>	1.003	0.968	1.040	0.85
PC3 <sub>aesthetic</sub>	0.973	0.918	1.031	PC3 <sub>aesthetic</sub>	0.979	0.933	1.028	0.40
PC4 <sub>aesthetic</sub>	0.993	0.929	1.061	PC4 <sub>aesthetic</sub>	1.007	0.950	1.067	0.81
COPD prevalence	1.039	0.840	1.284	COPD prevalence	1.040	0.841	1.285	0.72
Education <sup>1</sup>	0.976	0.923	1.033	Education <sup>1</sup>	0.986	0.931	1.043	0.62
Income <sup>2</sup>	1.000	0.999	1.002	Income <sup>2</sup>	1.000	0.999	1.001	0.80
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.050	0.012	Unstructured RE		0.008	0.001	
Spatially structured RE		0.080	0.011					
Burn in period = 60,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.040	0.994	1.089	Intercept	1.021	0.978	1.067	0.35
PC1 <sub>aesthetic</sub>	1.018	0.968	1.069	PC1 <sub>aesthetic</sub>	1.012	0.967	1.059	0.60
PC2 <sub>aesthetic</sub> Quintile 1	Reference category			PC2 <sub>aesthetic</sub> Quintile 1	Reference category			0.75 <sub>3</sub>
PC2 <sub>aesthetic</sub> Quintile 2	0.969	0.918	1.023	PC2 <sub>aesthetic</sub> Quintile 2	0.979	0.926	1.034	0.45
PC2 <sub>aesthetic</sub> Quintile 3	0.964	0.910	1.022	PC2 <sub>aesthetic</sub> Quintile 3	0.977	0.925	1.032	0.41
PC2 <sub>aesthetic</sub> Quintile 4	0.951	0.897	1.009	PC2 <sub>aesthetic</sub> Quintile 4	0.964	0.914	1.018	0.19
PC2 <sub>aesthetic</sub> Quintile 5	0.951	0.893	1.011	PC2 <sub>aesthetic</sub> Quintile 5	0.973	0.923	1.027	0.32
PC3 <sub>aesthetic</sub>	0.944	0.878	1.016	PC3 <sub>aesthetic</sub>	1.002	0.937	1.072	0.95
COPD prevalence	1.031	0.835	1.273	COPD prevalence	1.042	0.884	1.288	0.70
Education <sup>1</sup>	0.979	0.926	1.036	Education <sup>1</sup>	0.985	0.931	1.042	0.60
Income <sup>2</sup>	1.000	0.999	1.002	Income <sup>2</sup>	1.000	0.999	1.001	0.81
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.050	0.012	Unstructured RE		0.008	0.001	
Spatially structured RE		0.079	0.011					
Burn in period = 60,000 iterations								
Number of observations = 1344								

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars). <sup>3</sup>Overall p value for categorized variable based on type III likelihood ratio test.

In the frequentist models examining associations between aesthetic objective PC scores and hypertension prevalence for 2006-2009 in public water supplies, an increased PC<sub>3aesthetic</sub> score was associated with decreased hypertension prevalence for 2006 only (RR=0.943, 95% CI 0.891-0.999). No other significant associations between PC<sub>aesthetic</sub> scores in public or private water supplies and hypertension were identified for 2006-2009 (Table 6.10).

**Table 6.10** Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and hypertension prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	<b>2009</b>				<b>2008</b>				<b>2007</b>				<b>2006</b>			
	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	1.000	0.976	1.025	0.99	0.995	0.971	1.021	0.72	0.993	0.968	1.020	0.63	0.991	0.964	1.018	0.50
PC1 <sub>aesthetic</sub>	1.006	0.959	1.056	0.80	1.013	0.965	1.063	0.61	1.013	0.962	1.067	0.62	1.011	0.958	1.068	0.68
PC2 <sub>aesthetic</sub>	1.013	0.976	1.052	0.50	1.024	0.986	1.064	0.22	1.024	0.984	1.066	0.24	1.012	0.970	1.056	0.58
PC3 <sub>aesthetic</sub>	0.978	0.930	1.029	0.39	0.975	0.927	1.026	0.33	0.964	0.913	1.017	0.18	<b>0.943</b>	<b>0.891</b>	<b>0.999</b>	<b>0.045</b>
PC4 <sub>aesthetic</sub>	1.013	0.954	1.077	0.67	1.014	0.954	1.078	0.66	1.022	0.958	1.091	0.51	1.039	0.970	1.113	0.28
COPD prevalence	0.967	0.775	1.208	0.77	0.974	0.766	1.238	0.83	0.914	0.708	1.181	0.49	0.850	0.639	1.131	0.26
Education <sup>1</sup>	0.999	0.942	1.058	0.97	1.010	0.953	1.071	0.74	1.024	0.964	1.087	0.45	1.038	0.976	1.105	0.24
Income <sup>2</sup>	1.000	0.999	1.001	0.76	1.000	0.999	1.001	0.93	1.000	0.999	1.001	0.92	1.000	0.998	1.001	0.76
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.009		0.001		0.009		0.001		0.010		0.002		0.012		0.002	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	1.002	0.977	1.028	0.86	0.997	0.972	1.024	0.84	0.995	0.968	1.022	0.72	0.992	0.965	1.021	0.59
PC1 <sub>aesthetic</sub>	1.009	0.966	1.054	0.68	1.011	0.968	1.057	0.62	1.020	0.973	1.068	0.42	1.029	0.979	1.081	0.26
PC2 <sub>aesthetic</sub>	0.971	0.936	1.008	0.12	0.980	0.943	1.017	0.28	0.981	0.943	1.021	0.34	0.980	0.939	1.022	0.35
PC3 <sub>aesthetic</sub>	1.000	0.934	1.071	0.99	0.998	0.931	1.070	0.95	1.000	0.929	1.077	0.99	0.995	0.920	1.076	0.89
COPD prevalence	0.976	0.782	1.218	0.83	0.989	0.779	1.257	0.93	0.933	0.722	1.205	0.59	0.871	0.654	1.158	0.34
Education <sup>1</sup>	0.997	0.941	1.056	0.92	1.007	0.950	1.068	0.81	1.020	0.960	1.083	0.52	1.033	0.971	1.100	0.30
Income <sup>2</sup>	1.000	0.998	1.001	0.66	1.000	0.999	1.001	0.82	1.000	0.999	1.001	0.84	1.000	0.998	1.001	0.76
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.009		0.001		0.009		0.001		0.010		0.002		0.012		0.002	

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).



#### 6.4.4 Ischemic heart disease and arsenic concentrations

Drinking water arsenic concentration in public water supplies or private wells was not associated with prevalence of ischemic heart disease in the spatial Bayesian model or the frequentist model for 2010 (Table 6.11). Similarly, no associations between arsenic concentration and IHD prevalence were identified in the frequentist models for 2006-2009 (Table 6.12).

**Table 6.11** Comparison of Bayesian and frequentist models for the association between arsenic concentrations in drinking water and ischemic heart disease prevalence for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.946	0.861	1.040	Intercept	0.916	0.837	1.003	0.06
Arsenic (µg/L)	1.000	0.928	1.075	Arsenic (µg/L)	1.032	0.964	1.104	0.36
COPD prevalence	1.672	1.087	2.569	COPD prevalence	1.683	1.093	2.592	0.02
Education <sup>1</sup>	0.924	0.820	1.043	Education <sup>1</sup>	0.929	0.823	1.047	0.23
Income <sup>2</sup>	0.999	0.996	1.001	Income <sup>2</sup>	0.999	0.996	1.001	0.24
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.141	0.018	Unstructured RE		0.032	0.005	
Spatially structured RE		0.117	0.023					
Burn in period = 140,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.930	0.854	1.013	Intercept	0.912	0.838	0.992	0.03
Arsenic (µg/L)	1.011	0.969	1.055	Arsenic (µg/L)	1.025	0.984	1.068	0.24
COPD prevalence	1.681	1.095	2.584	COPD prevalence	1.692	1.098	2.605	0.02
Education <sup>1</sup>	0.923	0.818	1.042	Education <sup>1</sup>	0.927	0.822	1.046	0.22
Income <sup>2</sup>	0.999	0.996	1.001	Income <sup>2</sup>	0.999	0.996	1.001	0.25
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.143	0.017	Unstructured RE		0.032	0.005	
Spatially structured RE		0.113	0.023					
Burn in period = 120,000 iterations								
Number of observations = 1344								

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

**Table 6.12** Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and ischemic heart disease prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006			
	95% CI				95% CI				95% CI				95% CI			
	Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper
Intercept	0.095	0.825	0.993	0.04	0.931	0.848	1.022	0.13	0.957	0.870	1.052	0.36	0.990	0.902	1.087	0.83
Arsenic (µg/L)	1.031	0.963	1.105	0.38	1.021	0.953	1.094	0.55	0.998	0.931	1.070	0.95	0.985	0.921	1.054	0.66
COPD prevalence	2.154	1.367	3.394	0.001	2.325	1.421	3.805	<0.001	2.568	1.516	4.352	<0.001	2.505	1.394	4.502	0.002
Education <sup>1</sup>	0.907	0.801	1.027	0.12	0.865	0.762	0.982	0.03	0.846	0.742	0.964	0.01	0.837	0.731	0.959	0.01
Income <sup>2</sup>	0.998	0.995	1.001	0.12	0.998	0.995	1.001	0.14	0.997	0.994	1.000	0.03	0.997	0.995	1.000	0.07
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.033		0.005		0.033		0.006		0.032		0.006		0.027		0.005	
<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.894	0.820	0.974	0.01	0.926	0.849	1.010	0.08	0.951	0.871	1.039	0.27	0.974	0.893	1.063	0.56
Arsenic (µg/L)	1.030	0.988	1.074	0.17	1.018	0.976	1.062	0.40	1.002	0.961	1.046	0.91	1.000	0.960	1.042	0.99
COPD prevalence	2.168	1.376	3.416	<0.001	2.329	1.423	3.811	<0.001	2.569	1.516	4.353	<0.001	2.505	1.394	4.502	0.002
Education <sup>1</sup>	0.906	0.800	1.025	0.12	0.864	0.761	0.981	0.02	0.846	0.742	0.964	0.01	0.838	0.731	0.959	0.01
Income <sup>2</sup>	0.998	0.995	1.001	0.12	0.998	0.995	1.001	0.15	0.997	0.994	1.000	0.03	0.997	0.994	1.000	0.06
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.033		0.005		0.033		0.006		0.032		0.006		0.027		0.005	

Estimates are adjusted for age and sex status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Total average income in thousands of dollars

#### *6.4.5 Ischemic heart disease and health standards PC scores*

No associations between the health standards PC scores in public or private water supplies and IHD prevalence were evident in the Bayesian models for 2010 (Table 6.13). However, the frequentist model suggested that an increase in PC1<sub>health</sub> score was associated with a decrease in IHD prevalence for 2010 for both public water supplies (RR=0.832 95% CI 0.749-0.924,  $p<0.001$ ) and private wells (RR=0.885, 95% CI 0.805-0.973,  $p=0.01$ ). PC2<sub>health</sub> and PC3<sub>health</sub> scores were not associated with IHD prevalence in the 2010 frequentist model for either type of water supply (Table 6.13).

**Table 6.13** Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and ischemic heart disease prevalence for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.953	0.903	1.006	Intercept	0.959	0.908	1.013	0.14
PC1 <sub>health</sub>	0.921	0.807	1.053	PC1 <sub>health</sub>	<b>0.832</b>	<b>0.749</b>	<b>0.924</b>	<b>&lt;0.001</b>
PC2 <sub>health</sub>	0.988	0.891	1.094	PC2 <sub>health</sub>	1.028	0.941	1.122	0.54
PC3 <sub>health</sub>	0.967	0.877	1.065	PC3 <sub>health</sub>	0.976	0.895	1.064	0.58
COPD prevalence	1.677	1.088	2.570	COPD prevalence	1.695	1.103	2.606	0.02
Education <sup>1</sup>	0.922	0.818	1.039	Education <sup>1</sup>	0.923	0.818	1.041	0.19
Income <sup>2</sup>	0.999	0.996	1.001	Income <sup>2</sup>	0.998	0.996	1.001	0.21
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.145	0.019	Unstructured RE		0.030	0.005	
Spatially structured RE		0.104	0.029					
Burn in period = 180,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.942	0.891	0.997	Intercept	0.944	0.892	1.000	0.05
PC1 <sub>health</sub>	0.926	0.836	1.027	PC1 <sub>health</sub>	<b>0.885</b>	<b>0.805</b>	<b>0.973</b>	<b>0.01</b>
PC2 <sub>health</sub>	1.045	0.961	1.135	PC2 <sub>health</sub>	1.054	0.979	1.134	0.16
PC3 <sub>health</sub>	1.075	0.942	1.226	PC3 <sub>health</sub>	1.106	0.976	1.254	0.11
COPD prevalence	1.680	1.092	2.586	COPD prevalence	1.685	1.096	2.590	0.02
Education <sup>1</sup>	0.922	0.817	1.041	Education <sup>1</sup>	0.918	0.814	1.036	0.17
Income <sup>2</sup>	0.998	0.996	1.001	Income <sup>2</sup>	0.998	0.996	1.001	0.18
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.143	0.017	Unstructured RE		0.028	0.005	
Spatially structured RE		0.100	0.024					
Burn in period = 100,000 iterations								
Number of observations = 1344								

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

In the frequentist models for 2006-2009, and increase in PC1<sub>health</sub> score for public water supplies was significantly associated with decreased IHD prevalence in 2007 (RR=0.873, 95% CI 0.782-0.974), 2008 (RR=0.858, 95% CI 0.770-0.956), and 2009 (RR=0.833, 95% CI 0.748-0.927), but not in 2006 (Table 6.14). For private wells, increasing PC1<sub>health</sub> score was associated with decreased IHD prevalence in 2008 (RR=0.904, 95% CI 0.819-0.997, p=0.04) and 2009

(RR=0.878, 95% CI 0.798-0.967,  $p=0.009$ ) but not in 2006 or 2007 (Table 6.14). The second and third PC<sub>health</sub> scores were not significantly associated with IHD prevalence in 2006-2009 for public or private water supplies. Categorization of PC1<sub>health</sub> into quintiles demonstrated that the relationship between the quintiles of PC1<sub>health</sub> for public water supplies and IHD prevalence was not strictly monotonic, but the prevalence of IHD was significantly lower with higher quintiles of PC1<sub>health</sub> compared to the first quintile (not shown).

**Table 6.14** Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and ischemic heart disease prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006			
	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.948	0.896	1.002	0.06	0.963	0.909	1.020	0.20	0.968	0.913	1.026	0.27	0.987	0.930	1.047	0.66
PC1 <sub>health</sub>	<b>0.833</b>	<b>0.748</b>	<b>0.927</b>	<b>&lt;0.001</b>	<b>0.858</b>	<b>0.770</b>	<b>0.956</b>	<b>0.006</b>	<b>0.873</b>	<b>0.782</b>	<b>0.974</b>	<b>0.02</b>	0.910	0.819	1.012	0.08
PC2 <sub>health</sub>	1.016	0.929	1.112	0.72	1.001	0.914	1.097	0.98	0.983	0.897	1.077	0.71	0.963	0.881	1.053	0.41
PC3 <sub>health</sub>	0.970	0.888	1.059	0.50	0.974	0.891	1.066	0.57	0.946	0.864	1.036	0.23	0.941	0.862	1.027	0.17
COPD prevalence	2.170	1.380	3.414	<0.001	2.349	1.438	3.840	<0.001	2.620	1.548	4.435	<0.001	2.541	1.415	4.561	0.002
Education <sup>1</sup>	0.900	0.795	1.019	0.10	0.858	0.756	0.974	0.02	0.837	0.734	0.954	0.008	0.828	0.723	0.948	0.007
Income <sup>2</sup>	0.998	0.995	1.000	0.10	0.998	0.995	1.001	0.13	0.997	0.994	1.000	0.03	0.997	0.995	1.000	0.06
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.030		0.005		0.031		0.005		0.030		0.005		0.026		0.005	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.931	0.879	0.987	0.02	0.948	0.893	1.006	0.08	0.955	0.898	1.015	0.14	0.974	0.916	1.036	0.40
PC1 <sub>health</sub>	<b>0.878</b>	<b>0.798</b>	<b>0.967</b>	<b>0.009</b>	<b>0.904</b>	<b>0.819</b>	<b>0.997</b>	<b>0.04</b>	0.910	0.823	1.006	0.07	0.935	0.847	1.031	0.18
PC2 <sub>health</sub>	1.044	0.968	1.125	0.26	1.034	0.958	1.116	0.39	1.027	0.950	1.110	0.50	1.006	0.933	1.085	0.88
PC3 <sub>health</sub>	1.109	0.976	1.259	0.11	1.094	0.961	1.245	0.17	1.049	0.919	1.197	0.48	1.023	0.899	1.164	0.73
COPD prevalence	2.155	1.370	3.389	<0.001	2.340	1.433	3.823	<0.001	2.598	1.535	4.396	<0.001	2.520	1.403	4.527	<0.001
Education <sup>1</sup>	0.897	0.793	1.016	0.09	0.857	0.755	0.972	0.02	0.839	0.736	0.956	0.008	0.834	0.728	0.955	0.009
Income <sup>2</sup>	0.998	0.995	1.000	0.08	0.998	0.995	1.000	0.11	0.997	0.994	1.000	0.02	0.997	0.994	1.000	0.05
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.029		0.005		0.030		0.005		0.030		0.005		0.026		0.005	

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

#### 6.4.6 Ischemic heart disease and aesthetic objective PC scores

In the Bayesian model examining associations between aesthetic objective PC scores from public water supplies and IHD in 2010, increasing scores for PC2<sub>aesthetic</sub> were associated with a decrease in the prevalence of IHD (RR=0.913, 95% CrI 0.836-0.994) with a posterior probability that  $RR < 1$  of 98.2%. However, when PC2<sub>aesthetic</sub> was categorized into quintiles, there were no differences in IHD prevalence between the lowest and higher quintiles of PC2<sub>aesthetic</sub> where 1 was not included in the 95% CrI (Table 6.15). The relationship between PC2<sub>health</sub> and IHD prevalence did not appear to be completely linear or monotonic. The remaining aesthetic objective PC scores were not associated with IHD prevalence (Table 6.15).

When the PC<sub>aesthetic</sub> scores from public water supplies were assessed as continuous terms in the frequentist model for IHD in 2010, an increase in PC2<sub>aesthetic</sub> score was associated with a decrease in IHD prevalence (RR= 0.908, 95% CI 0.846-0.975,  $p=0.008$ ), as was an increase in PC1<sub>aesthetic</sub> score (RR= 0.877, 0.801-0.960,  $p=0.005$ ). However, when PC2<sub>aesthetic</sub> was categorized for direct comparison to the final Bayesian model, there were no significant differences in IHD prevalence between the lowest and higher PC2<sub>aesthetic</sub> quintiles (Table 6.15). However, the type III test of fixed effects for the categorized PC2<sub>aesthetic</sub> was significant ( $p=0.03$ ). Other pairwise differences were significant only for the fifth quintile compared to the third (RR=0.869, 95% CI 0.783-0.964,  $p=0.01$ ), the fifth quintile compared to the second (RR= 0.895, 95% CI 0.807-0.993,  $p=0.04$ ), and the fourth quintile compared to the second (RR= 0.887, 95% CI 0.801-0.981,  $p=0.02$ ) and, therefore, the pattern of differences was not consistent with a monotonic exposure-response relationship. The remaining PC<sub>aesthetic</sub> scores were not significantly associated with IHD prevalence (Table 6.15)

Aesthetic objective PC scores for private well water were not associated with IHD prevalence for 2010 in the Bayesian model (Table 6.15). However, in the frequentist model, an increase in the score for PC1<sub>aesthetic</sub> was associated with a decrease in IHD prevalence (RR=0.868, 95% CI 0.800-0.941), as was the score for PC2<sub>aesthetic</sub> (RR= 0.923, 95% CI 0.860-0.990). The third PC<sub>aesthetic</sub> was not significantly associated with IHD prevalence (Table 16.5).

**Table 6.15** Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and ischemic heart disease prevalence for 2010 in rural southern Saskatchewan.

Bayesian model				Frequentist model				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.966	0.880	1.062	Intercept	0.968	0.887	1.057	0.47
PC1 <sub>aesthetic</sub>	0.956	0.863	1.060	PC1 <sub>aesthetic</sub>	0.916	0.838	1.002	0.054
PC2 <sub>aesthetic</sub> Quintile 1	reference category			PC2 <sub>aesthetic</sub> Quintile 1	reference category			0.03 <sup>3</sup>
PC2 <sub>aesthetic</sub> Quintile 2	1.004	0.899	1.121	PC2 <sub>aesthetic</sub> Quintile 2	1.019	0.916	1.133	0.73
PC2 <sub>aesthetic</sub> Quintile 3	1.052	0.937	1.178	PC2 <sub>aesthetic</sub> Quintile 3	1.050	0.942	1.171	0.38
PC2 <sub>aesthetic</sub> Quintile 4	0.944	0.835	1.065	PC2 <sub>aesthetic</sub> Quintile 4	0.931	0.836	1.037	0.20
PC2 <sub>aesthetic</sub> Quintile 5	0.889	0.777	1.010	PC2 <sub>aesthetic</sub> Quintile 5	0.912	0.818	1.018	0.10
PC3 <sub>aesthetic</sub>	0.980	0.871	1.099	PC3 <sub>aesthetic</sub>	1.060	0.963	1.167	0.24
PC4 <sub>aesthetic</sub>	1.085	0.957	1.231	PC4 <sub>aesthetic</sub>	1.089	0.972	1.221	0.14
COPD prevalence	1.694	1.097	2.615	COPD prevalence	1.733	1.127	2.665	0.01
Education <sup>1</sup>	0.928	0.821	1.047	Education <sup>1</sup>	0.921	0.817	1.039	0.18
Income <sup>2</sup>	0.998	0.996	1.001	Income <sup>2</sup>	0.999	0.996	1.001	0.25
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.134	0.020	Unstructured RE		0.027	0.005	
Spatially structured RE		0.114	0.027					

Burn in period = 120,000 iterations

<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.960	0.909	1.013	Intercept	0.966	0.914	1.021	0.22
PC1 <sub>aesthetic</sub>	0.913	0.833	1.001	PC1 <sub>aesthetic</sub>	<b>0.868</b>	<b>0.800</b>	<b>0.941</b>	<b>&lt;0.001</b>
PC2 <sub>aesthetic</sub>	0.931	0.862	1.006	PC2 <sub>aesthetic</sub>	<b>0.923</b>	<b>0.860</b>	<b>0.990</b>	<b>0.03</b>
PC3 <sub>aesthetic</sub>	0.983	0.855	1.128	PC3 <sub>aesthetic</sub>	1.010	0.887	1.151	0.88
COPD prevalence	1.693	1.104	2.592	COPD prevalence	1.727	1.124	2.654	0.01
Education <sup>1</sup>	0.918	0.813	1.035	Education <sup>1</sup>	0.912	0.809	1.029	0.14
Income <sup>2</sup>	0.998	0.996	1.001	Income <sup>2</sup>	0.998	0.996	1.001	0.19
Random effects		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.143	0.017	Unstructured RE		0.028	0.005	
Spatially structured RE		0.100	0.026					

Burn in period = 80,000 iterations

Number of observations = 1344

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars). <sup>3</sup>Overall p value for categorized variable based on type III likelihood ratio test.



In the frequentist GLMM for 2006-2009, an increase in the  $PC1_{\text{aesthetic}}$  score in public water supplies was associated with a decrease in IHD prevalence for 2006 (RR=0.878, 95% CI 0.802-0.961), 2007 (RR=0.886, 95% CI 0.806-0.974), 2008 (RR=0.887, 95% CI 0.808-0.974), and 2009 (RR=0.877, 95% CI 0.800-0.962). The relationship between the quintiles of  $PC1_{\text{aesthetic}}$  from public water supplies and IHD prevalence was not clearly monotonic and similar patterns were evident for 2006-2010 (results not shown). An increase in  $PC2_{\text{aesthetic}}$  scores in public water supplies was also associated with decreased IHD prevalence for 2006 (RR=0.905, 95% CI 0.843-0.971), 2007 (RR=0.912, 95% CI 0.847-0.982), 2008 (RR=0.917, 95% CI 0.852-0.987), and 2009 (RR=0.910, 95% CI 0.847-0.978). The relationship between quintiles of  $PC2_{\text{aesthetic}}$  from public water supplies and IHD prevalence for 2006-2009 was non-monotonic (results not shown) and similar to that demonstrated in 2010 (Table 6.16). The third and fourth  $PC_{\text{aesthetic}}$  scores were not significantly associated with IHD prevalence in any year from 2006-2009 (Table 6.16).

An increase in the first  $PC_{\text{aesthetic}}$  score in private water supplies was associated with a decrease in IHD prevalence in 2007 (RR=0.903, 95% CI 0.829-0.983), 2008 (RR=0.895, 95% CI 0.822-0.974), and 2009 (RR=0.878, 95% CI 0.808-0.955), but not in 2006 (Table 6.16). The relationship between  $PC1_{\text{aesthetic}}$  from private wells and IHD prevalence was not clearly monotonic and similar for 2007-2010 (results not shown). An increase in  $PC2_{\text{aesthetic}}$  score in private wells was associated with a decrease in IHD prevalence in 2009 (RR=0.926, 95% CI 0.861-0.995) but not in any other year (Table 6.16). The third  $PC_{\text{aesthetic}}$  was not significantly associated with IHD prevalence in any year from 2006-2009 (Table 6.16).

**Table 6.16** Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and ischemic heart disease prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006			
	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.942	0.892	0.995	0.03	0.958	0.905	1.013	0.13	0.959	0.906	1.016	0.16	0.977	0.923	1.035	0.44
PC1 <sub>aesthetic</sub>	<b>0.877</b>	<b>0.800</b>	<b>0.962</b>	<b>0.006</b>	<b>0.887</b>	<b>0.808</b>	<b>0.974</b>	<b>0.01</b>	<b>0.886</b>	<b>0.806</b>	<b>0.974</b>	<b>0.01</b>	<b>0.878</b>	<b>0.802</b>	<b>0.961</b>	<b>0.005</b>
PC2 <sub>aesthetic</sub>	<b>0.910</b>	<b>0.847</b>	<b>0.978</b>	<b>0.01</b>	<b>0.917</b>	<b>0.852</b>	<b>0.987</b>	<b>0.02</b>	<b>0.912</b>	<b>0.847</b>	<b>0.982</b>	<b>0.01</b>	<b>0.905</b>	<b>0.843</b>	<b>0.971</b>	<b>0.005</b>
PC3 <sub>aesthetic</sub>	1.065	0.967	1.172	0.20	1.052	0.954	1.160	0.31	1.019	0.923	1.125	0.71	0.976	0.888	1.073	0.62
PC4 <sub>aesthetic</sub>	1.077	0.959	1.209	0.21	1.084	0.964	1.220	0.18	1.086	0.964	1.223	0.17	1.108	0.989	1.241	0.08
COPD prevalence	2.218	1.410	3.492	<0.001	2.369	1.449	3.872	<0.001	2.639	1.559	4.468	<0.001	2.592	1.445	4.650	0.001
Education <sup>1</sup>	0.899	0.795	1.018	0.09	0.860	0.758	0.976	0.02	0.843	0.740	0.961	0.01	0.838	0.732	0.959	0.01
Income <sup>2</sup>	0.998	0.995	1.000	0.07	0.998	0.995	1.000	0.09	0.997	0.994	0.999	0.02	0.997	0.994	0.999	0.02
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.029		0.005		0.029		0.005		0.029		0.005		0.023		0.005	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.952	0.900	1.007	0.09	0.967	0.913	1.025	0.26	0.971	0.915	1.030	0.33	0.987	0.930	1.047	0.66
PC1 <sub>aesthetic</sub>	<b>0.878</b>	<b>0.808</b>	<b>0.955</b>	<b>0.002</b>	<b>0.895</b>	<b>0.822</b>	<b>0.974</b>	<b>0.01</b>	<b>0.903</b>	<b>0.829</b>	<b>0.983</b>	<b>0.02</b>	0.931	0.857	1.012	0.09
PC2 <sub>aesthetic</sub>	<b>0.926</b>	<b>0.861</b>	<b>0.995</b>	<b>0.04</b>	0.938	0.871	1.009	0.09	0.929	0.862	1.000	0.051	0.934	0.869	1.004	0.06
PC3 <sub>aesthetic</sub>	1.020	0.893	1.166	0.77	1.005	0.878	1.151	0.94	0.988	0.862	1.133	0.87	0.994	0.871	1.136	0.93
COPD prevalence	2.206	1.403	3.471	<0.001	2.379	1.456	3.886	<0.001	2.628	1.554	4.446	<0.001	2.567	1.430	4.609	0.002
Education <sup>1</sup>	0.892	0.788	1.010	0.07	0.853	0.751	0.968	0.01	0.835	0.733	0.951	0.007	0.828	0.723	0.949	0.007
Income <sup>2</sup>	0.998	0.995	1.000	0.09	0.998	0.995	1.001	0.11	0.997	0.994	0.9995	0.02	0.997	0.994	0.9999	0.04
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.030		0.005		0.030		0.005		0.030		0.005		0.025		0.005	

Estimates are adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

#### 6.4.7 Stroke and arsenic concentrations

There were no associations identified between arsenic concentration in drinking water and stroke prevalence for 2010 in the Bayesian or frequentist models for public water supplies or private wells (Table 6.17).

**Table 6.17** Comparison of Bayesian and frequentist models for the association between arsenic concentrations in drinking water and stroke prevalence for 2010 in rural southern Saskatchewan.

Bayesian models					Frequentist models				
<i>Public water supplies</i>									
Effect	RR	95% CrI			Effect	RR	95% CI		p
		lower	upper				lower	upper	
Intercept	1.005	0.909	1.110		Intercept	0.997	0.907	1.095	0.95
Arsenic (µg/L)	0.975	0.906	1.047		Arsenic (µg/L)	0.980	0.918	1.046	0.54
COPD prevalence	2.301	1.296	4.063		COPD prevalence	2.304	1.307	4.062	0.004
Education <sup>1</sup>	0.837	0.717	0.979		Education <sup>1</sup>	0.841	0.721	0.983	0.03
Income <sup>2</sup>	0.998	0.994	1.001		Income <sup>2</sup>	0.998	0.994	1.001	0.15
Random effects		SD	SE		Random effect		Variance	SE	
Unstructured RE		0.081	0.029		Unstructured RE		0.015	0.004	
Spatially structured RE		0.091	0.031						
Burn in period = 140,000 iterations									
<i>Private wells</i>									
Effect	RR	95% CrI			Effect	RR	95% CI		p
		lower	upper				lower	upper	
Intercept	0.978	0.890	1.074		Intercept	0.970	0.887	1.061	0.51
Arsenic (µg/L)	1.000	0.957	1.044		Arsenic (µg/L)	1.003	0.964	1.044	0.87
COPD prevalence	2.316	1.325	4.080		COPD prevalence	2.298	1.303	4.054	0.004
Education <sup>1</sup>	0.837	0.717	0.977		Education <sup>1</sup>	0.843	0.722	0.984	0.03
Income <sup>2</sup>	0.998	0.994	1.001		Income <sup>2</sup>	0.997	0.994	1.001	0.14
Random effects		SD	SE		Random effect		Variance	SE	
Unstructured RE		0.085	0.029		Unstructured RE		0.015	0.004	
Spatially structured RE		0.087	0.030						
Burn in period = 140,000 iterations									
Number of observations = 1344									

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

For public water supplies, increased arsenic concentration was associated with a decrease in stroke prevalence in the frequentist model for 2007 only (RR= 0.927, 95% CI 0.862-0.998) (Table 6.18). For private water supplies, no associations between arsenic concentration and stroke prevalence were evident in the frequentist GLMMs for 2006-2009 (Table 6.18).

**Table 6.18** Summary of the results for frequentist GLMM examining the association between arsenic concentrations in drinking water and stroke prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006				
	95% CI				95% CI				95% CI				95% CI				
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	
Intercept	1.013	0.919	1.116	0.80	1.033	0.935	1.141	0.52	1.047	0.943	1.162	0.39	1.051	0.945	1.169	0.36	
Arsenic (µg/L)	0.963	0.901	1.030	0.28	0.945	0.882	1.012	0.11	<b>0.927 0.862 0.998 0.04</b>	0.940	0.875	1.010	0.09	0.940	0.875	1.010	0.09
COPD prevalence	2.580	1.415	4.702	0.002	1.841	0.954	3.553	0.07	2.023	0.987	4.144	0.05	1.250	0.555	2.818	0.59	
Education <sup>1</sup>	0.832	0.709	0.976	0.02	0.910	0.772	1.072	0.26	0.911	0.767	1.083	0.29	0.974	0.813	1.168	0.78	
Income <sup>2</sup>	0.998	0.995	1.002	0.27	0.999	0.996	1.003	0.73	0.998	0.994	1.002	0.26	0.998	0.994	1.001	0.22	
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE		
	0.016		0.004		0.015		0.005		0.017		0.005		0.011		0.005		
<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI				
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	
Intercept	0.981	0.894	1.076	0.68	1.018	0.926	1.119	0.71	1.013	0.917	1.119	0.80	1.007	0.910	1.115	0.89	
Arsenic (µg/L)	0.995	0.955	1.037	0.82	0.971	0.930	1.013	0.17	0.970	0.927	1.014	0.18	0.985	0.942	1.029	0.50	
COPD prevalence	2.563	1.405	4.675	0.002	1.818	0.943	3.507	0.07	1.992	0.972	4.082	0.06	1.236	0.547	2.791	0.61	
Education <sup>1</sup>	0.835	0.711	0.979	0.03	0.915	0.777	1.078	0.29	0.918	0.773	1.090	0.33	0.981	0.818	1.176	0.83	
Income <sup>2</sup>	0.998	0.995	1.001	0.26	0.999	0.996	1.003	0.70	0.998	0.994	1.002	0.24	0.998	0.994	1.001	0.21	
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE		
	0.016		0.005		0.016		0.005		0.017		0.005		0.013		0.005		

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Total average income in thousands of dollars

#### *6.4.8 Stroke and health standards PC scores*

In the Bayesian model examining associations between health standards PC scores for public water supplies and stroke prevalence, an increase in  $PC1_{\text{health}}$  was associated with an increase in stroke prevalence for 2010 ( $RR=1.133$ , 95% CrI 1.003-1.281) when all PC scores were evaluated as continuous variables. The posterior probability that  $RR>1$  for  $PC1_{\text{health}}$  was 97.8%. When the  $PC1_{\text{health}}$  scores were categorized into quintiles with the first quintile as the reference category, there were no differences in stroke prevalence between the first and higher quintiles for  $PC1_{\text{health}}$  where 1 was not included in the CrI (Table 6.19). In the frequentist model for 2010 an increase in  $PC1_{\text{health}}$  scores was not significantly associated with an increase in stroke prevalence when evaluated as a continuous variable ( $RR=1.103$ , 95% CI 0.999-1.218,  $p=0.053$ ), but  $PC2_{\text{health}}$  was associated with decreased stroke prevalence ( $RR=0.916$ , 95% CI 0.841-0.999,  $p=0.048$ ).  $PC1_{\text{health}}$  was evaluated as a categorized variable in a frequentist GLMM for direct comparison to the Bayesian model. There was no significant difference in stroke prevalence between the second through fifth quintiles compared to the first (Table 6.19), and the type III test of fixed effects for the categorized variable was not significant ( $p=0.30$ ).

For private wells, there were no associations between any of the health standards PC scores and stroke prevalence for 2010 in the Bayesian model (Table 6.19). For the frequentist model, an increase in the  $PC3_{\text{health}}$  score was associated with a decrease in stroke prevalence ( $RR=0.883$ , 95% CI 0.781-0.999).  $PC1_{\text{health}}$  and  $PC2_{\text{health}}$  were not associated with stroke prevalence for 2010 in the frequentist model (Table 6.19).

**Table 6.19** Comparison of Bayesian and frequentist models for the association between drinking water health standards principal component scores and stroke prevalence for 2010 in rural southern Saskatchewan.

Bayesian models				Frequentist models				
<i>Public water supplies</i>								
Effect	RR	95% CrI			RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.918	0.834	1.012	Intercept	0.921	0.842	1.008	0.08
PC1 <sub>health</sub> Quintile 1	Reference category			PC1 <sub>health</sub> Quintile 1	Reference category			0.30 <sup>3</sup>
PC1 <sub>health</sub> Quintile 2	1.080	0.970	1.203	PC1 <sub>health</sub> Quintile 2	1.088	0.981	1.206	0.11
PC1 <sub>health</sub> Quintile 3	1.092	0.977	1.222	PC1 <sub>health</sub> Quintile 3	1.086	0.979	1.206	0.12
PC1 <sub>health</sub> Quintile 4	1.050	0.934	1.177	PC1 <sub>health</sub> Quintile 4	1.047	0.941	1.165	0.40
PC1 <sub>health</sub> Quintile 5	1.115	0.992	1.253	PC1 <sub>health</sub> Quintile 5	1.106	0.999	1.224	0.053
PC2 <sub>health</sub>	0.917	0.826	1.019	PC2 <sub>health</sub>	0.916	0.837	1.003	0.06
PC3 <sub>health</sub>	1.012	0.919	1.115	PC3 <sub>health</sub>	1.015	0.931	1.106	0.73
COPD prevalence	2.212	1.256	3.900	COPD prevalence	2.252	1.277	3.970	0.005
Education <sup>1</sup>	0.843	0.721	0.987	Education <sup>1</sup>	0.842	0.720	0.983	0.03
Income <sup>2</sup>	0.998	0.995	1.001	Income <sup>2</sup>	0.998	0.994	1.001	0.21
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.090	0.029	Unstructured RE		0.014	0.004	
Spatially structured RE		0.081	0.032					
Burn in period = 100,000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI			RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.996	0.933	1.062	Intercept	0.995	0.932	1.062	0.88
PC1 <sub>health</sub>	1.067	0.961	1.184	PC1 <sub>health</sub>	1.058	0.963	1.162	0.24
PC2 <sub>health</sub>	1.070	0.984	1.166	PC2 <sub>health</sub>	1.055	0.982	1.133	0.14
PC3 <sub>health</sub>	0.892	0.783	1.017	PC3 <sub>health</sub>	<b>0.883</b>	<b>0.781</b>	<b>0.999</b>	<b>0.048</b>
COPD prevalence	2.463	1.386	4.345	COPD prevalence	2.478	1.403	4.378	0.002
Education <sup>1</sup>	0.827	0.708	0.967	Education <sup>1</sup>	0.831	0.711	0.970	0.02
Income <sup>2</sup>	0.998	0.994	1.001	Income <sup>2</sup>	0.997	0.994	1.001	0.13
Random effects		SD	SE	Random effect		Variance	SE	
Unstructured RE		0.076	0.029	Unstructured RE		0.013	0.004	
Spatially structured RE		0.090	0.029					
Burn in period = 40,000 iterations								
Number of observations = 1344								

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars). <sup>3</sup>Overall p value for categorized variable based on type III likelihood ratio test.

In the frequentist GLMMs evaluating associations between health standards PC for public water supplies and stroke prevalence for 2006-2009, an increase in PC1<sub>health</sub> was associated with an increase in stroke prevalence (RR=1.130, 95% CI 1.021-1.250) for 2009 only. Similarly, an increase in PC2<sub>health</sub> was associated with a decrease in stroke prevalence (RR=0.894, 95% CI 0.818-0.976) for 2009 only. PC3<sub>health</sub> was not associated with stroke prevalence in any year (Table 6.20).

For the private well data, an increase in PC3<sub>health</sub> score was associated with a decrease in stroke prevalence in 2008 (RR=0.871, 95% CI 0.765-0.992) and 2009 (RR=0.873, 95% CI 0.769-0.991), but not in other years. PC1<sub>health</sub> and PC2<sub>health</sub> were not associated with stroke prevalence in any year (Table 6.20).

**Table 6.20** Summary of the results for frequentist GLMM examining the association between drinking water health standards principal component scores and stroke prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006			
	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.974	0.912	1.039	0.42	0.977	0.914	1.045	0.51	0.973	0.907	1.044	0.45	0.986	0.916	1.062	0.72
PC1 <sub>health</sub>	<b>1.130</b>	<b>1.021</b>	<b>1.250</b>	<b>0.02</b>	1.073	0.965	1.192	0.19	1.025	0.917	1.147	0.66	1.023	0.915	1.144	0.69
PC2 <sub>health</sub>	<b>0.894</b>	<b>0.818</b>	<b>0.976</b>	<b>0.01</b>	0.913	0.833	1.001	0.052	0.938	0.851	1.033	0.19	0.979	0.888	1.078	0.66
PC3 <sub>health</sub>	0.984	0.902	1.074	0.72	0.947	0.865	1.036	0.23	0.919	0.835	1.011	0.08	0.950	0.863	1.045	0.29
COPD prevalence	2.651	1.459	4.816	<0.001	1.848	0.959	3.562	0.07	2.063	1.007	4.226	0.048	1.263	0.559	2.851	0.57
Education <sup>1</sup>	0.824	0.703	0.966	0.02	0.903	0.766	1.064	0.22	0.903	0.759	1.073	0.25	0.973	0.810	1.168	0.77
Income <sup>2</sup>	0.998	0.995	1.002	0.34	1.000	0.996	1.003	0.81	0.998	0.994	1.002	0.27	0.998	0.994	1.001	0.22
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.013		0.004		0.014		0.005		0.016		0.005		0.012		0.005	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.997	0.932	1.066	0.93	0.994	0.928	1.065	0.87	0.983	0.914	1.057	0.64	0.996	0.924	1.075	0.92
PC1 <sub>health</sub>	1.055	0.957	1.163	0.28	1.058	0.957	1.169	0.27	1.055	0.949	1.173	0.32	1.068	0.961	1.186	0.22
PC2 <sub>health</sub>	1.076	0.999	1.158	0.052	1.054	0.977	1.138	0.17	1.059	0.977	1.147	0.16	1.032	0.953	1.117	0.44
PC3 <sub>health</sub>	<b>0.873</b>	<b>0.769</b>	<b>0.991</b>	<b>0.04</b>	<b>0.871</b>	<b>0.765</b>	<b>0.992</b>	<b>0.04</b>	0.911	0.794	1.046	0.19	0.920	0.802	1.054	0.23
COPD prevalence	2.807	1.536	5.127	<0.001	1.929	0.998	3.730	0.051	2.056	1.002	4.219	0.049	1.270	0.562	2.870	0.57
Education <sup>1</sup>	0.818	0.697	0.960	0.01	0.904	0.767	1.066	0.23	0.910	0.766	1.082	0.29	0.977	0.815	1.172	0.80
Income <sup>2</sup>	0.998	0.994	1.001	0.23	0.999	0.996	1.003	0.70	0.998	0.994	1.002	0.24	0.998	0.994	1.001	0.22
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.014		0.004		0.015		0.005		0.017		0.005		0.012		0.005	

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).



#### *6.4.9 Stroke and aesthetic objective PC scores*

An increase in  $PC3_{\text{aesthetic}}$  for public water supplies was associated with a decrease in stroke prevalence in 2010 (RR= 0.888, 95% CrI 0.803-0.984) when all PC scores were evaluated as continuous variables in the Bayesian model. The posterior probability that  $RR < 1$  was 98.8%.  $PC3_{\text{aesthetic}}$  was categorized into quintiles, and when the second through fifth quintiles were compared to the first, no differences in stroke prevalence were evident (Table 6.21). In the frequentist model for 2010, an increase in  $PC3_{\text{aesthetic}}$  was also associated with a decrease in stroke prevalence when evaluated as a continuous variable (RR=0.881, 95% CI 0.805-0.964,  $p=0.006$ ). In the frequentist model evaluating quintiles of  $PC3_{\text{aesthetic}}$  the type III test of fixed effects for the categorized form of  $PC3_{\text{aesthetic}}$  was not significant ( $p=0.12$ ).

There were no associations between aesthetic objective PC scores for private wells and stroke prevalence for 2010 in either the Bayesian or frequentist model (Table 6.21).

**Table 6.21** Comparison of Bayesian and frequentist models for the association between drinking water aesthetic objectives principal component scores and stroke prevalence for 2010 in rural southern Saskatchewan.

Bayesian models				Frequentist models				
<i>Public water supplies</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	1.033	0.945	1.128	Intercept	1.040	0.956	1.132	0.36
PC1 <sub>aesthetic</sub>	0.969	0.878	1.071	PC1 <sub>aesthetic</sub>	0.954	0.874	1.041	0.29
PC2 <sub>aesthetic</sub>	1.058	0.975	1.153	PC2 <sub>aesthetic</sub>	1.046	0.976	1.120	0.20
PC3 <sub>aesthetic</sub> Quintile 1	Reference category			PC3 <sub>aesthetic</sub> Quintile 1	Reference category			0.12 <sup>3</sup>
PC3 <sub>aesthetic</sub> Quintile 2	0.938	0.846	1.041	PC3 <sub>aesthetic</sub> Quintile 2	0.928	0.841	1.024	0.14
PC3 <sub>aesthetic</sub> Quintile 3	0.955	0.867	1.053	PC3 <sub>aesthetic</sub> Quintile 3	0.951	0.867	1.043	0.29
PC3 <sub>aesthetic</sub> Quintile 4	0.914	0.820	1.018	PC3 <sub>aesthetic</sub> Quintile 4	<b>0.903</b>	<b>0.815</b>	<b>0.999</b>	<b>0.049</b>
PC3 <sub>aesthetic</sub> Quintile 5	0.894	0.798	1.004	PC3 <sub>aesthetic</sub> Quintile 5	<b>0.881</b>	<b>0.797</b>	<b>0.973</b>	<b>0.01</b>
PC4 <sub>aesthetic</sub>	1.078	0.953	1.216	PC4 <sub>aesthetic</sub>	1.085	0.971	1.212	0.15
COPD prevalence	2.192	1.240	3.869	COPD prevalence	2.207	1.254	3.884	0.006
Education <sup>1</sup>	0.852	0.730	0.996	Education <sup>1</sup>	0.859	0.736	1.002	0.053
Income <sup>2</sup>	0.998	0.994	1.001	Income <sup>2</sup>	0.997	0.994	1.001	0.11
Random effects		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.091	0.027	Unstructured RE		0.012	0.004	
Spatially structured RE		0.069	0.031					
Burn in period = 80000 iterations								
<i>Private wells</i>								
Effect	RR	95% CrI		Effect	RR	95% CI		p
		lower	upper			lower	upper	
Intercept	0.983	0.921	1.047	Intercept	0.976	0.915	1.041	0.46
PC1 <sub>aesthetic</sub>	0.927	0.844	1.017	PC1 <sub>aesthetic</sub>	0.957	0.884	1.036	0.28
PC2 <sub>aesthetic</sub>	1.046	0.963	1.133	PC2 <sub>aesthetic</sub>	1.058	0.984	1.137	0.13
PC3 <sub>aesthetic</sub>	0.951	0.824	1.099	PC3 <sub>aesthetic</sub>	0.986	0.867	1.122	0.83
COPD prevalence	2.402	1.358	4.212	COPD prevalence	2.340	1.327	4.124	0.003
Education <sup>1</sup>	0.829	0.710	0.969	Education <sup>1</sup>	0.836	0.716	0.976	0.02
Income <sup>2</sup>	0.998	0.994	1.001	Income <sup>2</sup>	0.998	0.994	1.001	0.17
Random effects		SD	SE	Random effects		Variance	SE	
Unstructured RE		0.075	0.029	Unstructured RE		0.014	0.004	
Spatially structured RE		0.092	0.030					
Burn in period = 160000 iterations								
Number of observations = 1344								

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CrI = credible interval, CI = confidence interval, RR = risk ratio, SD = standard deviation, SE = standard error, PC = principal component, RE = random effect.

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars). <sup>3</sup>Overall p value for categorized variable based on type III likelihood ratio test.

In the frequentist models for 2006-2009, increased PC3<sub>aesthetic</sub> scores in public water supplies were associated with a decrease in stroke prevalence for 2008 (RR=0.877, 95% CI 0.797-0.966) and 2009 (RR=0.874, 95% CI 0.797-0.959), but not in 2006 or 2007 (Table 6.22). Increased PC4<sub>aesthetic</sub> scores in public supplies were also associated with an increased stroke prevalence for 2008 only (RR=1.132, 95% CI 1.007-1.273) (Table 6.22). No associations were demonstrated between any of the PC<sub>aesthetic</sub> scores in private wells and stroke prevalence for any year from 2006-2009 (Table 6.22).

**Table 6.22** Summary of the results for frequentist GLMM examining the association between drinking water aesthetic objectives principal component scores and stroke prevalence for 2006-2009 in rural southern Saskatchewan.

<i>Public water supplies</i>	2009				2008				2007				2006			
	95% CI				95% CI				95% CI				95% CI			
Effect	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p
Intercept	0.966	0.906	1.030	0.29	0.965	0.903	1.031	0.29	0.961	0.896	1.030	0.26	0.977	0.908	1.051	0.53
PC1 <sub>aesthetic</sub>	0.947	0.866	1.035	0.23	0.936	0.854	1.027	0.16	0.944	0.854	1.042	0.25	0.952	0.863	1.049	0.32
PC2 <sub>aesthetic</sub>	1.054	0.982	1.131	0.14	1.011	0.940	1.087	0.76	1.008	0.932	1.090	0.84	0.990	0.916	1.068	0.79
PC3 <sub>aesthetic</sub>	<b>0.874</b>	<b>0.797</b>	<b>0.959</b>	<b>0.005</b>	<b>0.877</b>	<b>0.797</b>	<b>0.966</b>	<b>0.008</b>	0.907	0.819	1.006	0.06	0.916	0.828	1.014	0.09
PC4 <sub>aesthetic</sub>	1.105	0.987	1.238	0.08	<b>1.132</b>	<b>1.007</b>	<b>1.273</b>	<b>0.04</b>	1.074	0.948	1.217	0.26	1.111	0.981	1.258	0.10
COPD prevalence	2.439	1.341	4.435	0.004	1.711	0.887	3.299	0.11	1.910	0.930	3.920	0.08	1.180	0.522	2.665	0.69
Education <sup>1</sup>	0.854	0.728	1.001	0.051	0.940	0.798	1.106	0.45	0.934	0.786	1.110	0.44	1.001	0.835	1.201	0.99
Income <sup>2</sup>	0.998	0.994	1.001	0.24	0.999	0.995	1.003	0.56	0.998	0.994	1.001	0.21	0.997	0.993	1.001	0.15
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.012		0.004		0.013		0.004		0.016		0.005		0.011		0.005	

<i>Private wells</i>	95% CI				95% CI				95% CI				95% CI			
	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	p	RR	lower	upper	P
Intercept	0.976	0.913	1.042	0.47	0.976	0.911	1.045	0.48	0.973	0.906	1.045	0.46	0.990	0.919	1.067	0.79
PC1 <sub>aesthetic</sub>	0.948	0.874	1.030	0.21	0.964	0.885	1.050	0.40	0.931	0.852	1.018	0.12	0.953	0.873	1.040	0.28
PC2 <sub>aesthetic</sub>	1.060	0.984	1.142	0.13	1.037	0.960	1.119	0.36	1.016	0.938	1.100	0.71	1.015	0.937	1.098	0.72
PC3 <sub>aesthetic</sub>	0.963	0.842	1.101	0.58	0.951	0.829	1.092	0.48	0.964	0.834	1.115	0.62	0.948	0.821	1.096	0.47
COPD prevalence	2.636	1.446	4.805	0.002	1.817	0.941	3.511	0.08	2.006	0.978	4.113	0.06	1.239	0.548	2.803	0.61
Education <sup>1</sup>	0.826	0.704	0.969	0.02	0.912	0.774	1.075	0.27	0.909	0.765	1.081	0.28	0.975	0.813	1.170	0.79
Income <sup>2</sup>	0.998	0.995	1.002	0.28	0.999	0.996	1.003	0.71	0.998	0.994	1.002	0.24	0.997	0.994	1.001	0.20
Random effects	Variance		SE		Variance		SE		Variance		SE		Variance		SE	
	0.015		0.004		0.016		0.005		0.017		0.005		0.012		0.005	

Estimates are adjusted for age and sex by stratification of the case counts and expected counts. CI = confidence interval, RR = risk ratio, PC = principal component, SE = standard error

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

#### *6.4.10 Summary of results*

The results from the Bayesian models for 2010 examining associations between each water quality risk factor (as continuous variables) and hypertension, IHD, and stroke prevalence were summarized (Table 6.23).

**Table 6.23** Summary of effect estimates from Bayesian analysis of associations between drinking water risk factors from public water supply and private well data and hypertension, ischemic heart disease, and stroke prevalence for 2010. Effect estimates for which the 95% credible interval did not include 1 are indicated with bold type.

Risk Factor	Hypertension		Ischemic Heart Disease		Stroke	
	<i>Public Supplies</i>	<i>Private Wells</i>	<i>Public Supplies</i>	<i>Private Wells</i>	<i>Public Supplies</i>	<i>Private Wells</i>
	RR (95% CrI)	RR (95% CrI)	RR (95% CrI)	RR (95% CrI)	RR (95% CrI)	RR (95% CrI)
<b>Arsenic</b>	0.97 (0.93-1.02)	0.99 (0.95-1.02)	1.00 (0.91-1.10)	1.01 (0.94-1.08)	0.99 (0.89-1.09)	1.03 (0.96-1.10)
<i>Health Standards</i>						
<b>PC1<sub>health</sub></b>	1.05 (0.98-1.12)	0.99 (0.94-1.05)	0.92 (0.81-1.05)	0.93 (0.84-1.03)	<b>1.13 (1.00-1.28)†</b>	1.07 (0.96-1.18)
<b>PC2<sub>health</sub></b>	0.97 (0.91-1.02)	0.97 (0.93-1.02)	0.99 (0.89-1.09)	1.05 (0.96-1.14)	0.92 (0.83-1.01)	1.07 (0.98-1.17)
<b>PC3<sub>health</sub></b>	0.96 (0.91-1.01)	0.99 (0.92-1.06)	0.97 (0.88-1.07)	1.08 (0.94-1.23)	1.01 (0.92-1.11)	0.89 (0.78-1.02)
<i>Aesthetic Objectives</i>						
<b>PC1<sub>aesthetic</sub></b>	1.04 (0.98-1.10)	1.01 (0.97-1.06)	0.92 (0.83-1.02)	0.91 (0.83-1.00)	0.96 (0.87-1.06)	0.93 (0.84-1.02)
<b>PC2<sub>aesthetic</sub></b>	0.98 (0.93-1.03)	<b>0.94 (0.90-0.98)</b>	<b>0.91 (0.84-0.99)*†</b>	0.93 (0.86-1.01)	1.06 (0.97-1.15)	1.05 (0.96-1.13)
<b>PC3<sub>aesthetic</sub></b>	0.97 (0.92-1.03)	0.94 (0.88-1.01)	1.00 (0.89-1.12)	0.98 (0.86-1.13)	<b>0.89 (0.80-0.98)</b>	0.95 (0.82-1.10)
<b>PC4<sub>aesthetic</sub></b>	0.99 (0.93-1.06)		1.08 (0.96-1.22)		1.08 (0.95-1.22)	
<b>1 (0 Pos, 1 Neg)</b>		<b>1 (0 Pos, 1 Neg)</b>		<b>2 (1 Pos, 1 Neg)</b>		

\*Effect estimate also statistically significant in frequentist models for at least 4 out of 5 years.

†Results for PC consistent with effects reported in literature for one or more individual contaminants contributing to that PC.

Estimates are from models with all risk factors as continuous variables and adjusted for age, sex, and First Nations status by stratification of the case counts and expected counts, and by prevalence of COPD, education level and income by inclusion as risk factors in model.

RR = risk ratio, CrI = credible interval, PC = principal component, Pos = positive association, Neg = negative association.

Because eight risk factors were evaluated for public water supplies and seven risk factors for private wells, a total of 15 variables were evaluated in the Bayesian models for 2010 for each outcome. Two of the four associations where 1 was not included in the credible interval involved PC2<sub>aesthetic</sub> with hypertension and then with IHD. Whereas PC1<sub>health</sub> and PC3<sub>aesthetic</sub> were associated with stroke. The chance of falsely identifying one association among all of the risk factors and either hypertension or IHD (Table 6.23) was 53.7%, while the chance of falsely identifying two associations between the various risk factors examined and stroke (Table 6.23) was 17.1%.

A summary of the total number of significant associations identified across 5 years of frequentist models examining associations between each of the water quality risk factors (as continuous variables) and hypertension, IHD, and stroke prevalence can be found in Table 6.24.

**Table 6.24** Total number of non-significant results and positive and negative associations between the water quality risk factors and hypertension, ischemic heart disease, and stroke prevalence from the frequentist GLMM analysis over 5 time periods.

	Hypertension						Ischemic Heart Disease						Stroke					
	Public water			Private well			Public water			Private well			Public water			Private well		
	NS	Pos	Neg	NS	Pos	Neg	NS	Pos	Neg	NS	Pos	Neg	NS	Pos	Neg	NS	Pos	Neg
<b>Arsenic</b>	1	0	4	5	0	0	5	0	0	5	0	0	4	0	1	5	0	0
<i>Health Standards</i>																		
<b>PC1<sub>health</sub></b>	3	2	0	5	0	0	1	0	4	2	0	3	4	1	0	5	0	0
<b>PC2<sub>health</sub></b>	0	0	5	5	0	0	5	0	0	5	0	0	3	0	2	5	0	0
<b>PC3<sub>health</sub></b>	2	0	3	5	0	0	5	0	0	5	0	0	5	0	0	2	0	3
<i>Aesthetic Objectives</i>																		
<b>PC1<sub>aesthetic</sub></b>	5	0	0	5	0	0	0	0	5	1	0	4	5	0	0	5	0	0
<b>PC2<sub>aesthetic</sub></b>	5	0	0	4	0	1	0	0	5*†	3	0	2	5	0	0	5	0	0
<b>PC3<sub>aesthetic</sub></b>	4	0	1	5	0	0	5	0	0	5	0	0	2	0	3*	5	0	0
<b>PC4<sub>aesthetic</sub></b>	5	0	0	-	-	-	5	0	0	-	-	-	4	1	0	-	-	-
<b>Total significant associations</b>	<b>16 (2 Pos, 14 Neg)</b>						<b>23 (0 Pos, 23 Neg)</b>						<b>11 (2 Pos, 9 Neg)</b>					

\*Effect estimate 95% credible interval did not include 1 in Bayesian analysis.

†Results for contaminants contributing to PC consistent with effects reported in literature.

Counts of associations based on models with all risk factors considered as continuous variables.

NS = not significant, Pos=positive association, Neg=negative association, PC=principal component

Across the frequentist models for each outcome over five years evaluated for 15 risk factors, 75 associations were tested. Across all the models for hypertension, a total of 16 significant

associations were identified; while 23 were identified across the models for IHD, and 11 were identified for the models for stroke (Table 6.24).

#### *6.4.11 Evaluation of model fit*

Evaluation of spatial autocorrelation in the residuals from the frequentist models revealed that there was statistically significant spatial autocorrelation in the unexplained variance for the models for each combination of risk factor group and outcome with the single exception of aesthetic objective PCs and stroke prevalence (Table 6.25). The residuals from models for each outcome without any fixed effects were also evaluated each outcome and exhibited significant spatial autocorrelation (Table 6.25). The DIC for the Bayesian models with spatial random effects was compared to and was lower than the DIC for models without spatial random effects for the four models (Table 6.23) for which associations were identified. The model investigating associations between aesthetic objective PCs from public water supplies (with PC2<sub>aesthetic</sub> categorized into quintiles) and IHD prevalence had a DIC of 5801 when the spatial random effect was included, but when the spatial random effect was not included the DIC was 5868. The model examining associations between health standards PCs from public water supplies and stroke prevalence had a DIC of 5124 when spatial random effects were included and 5128 when they were excluded. Similarly, the model for aesthetic objectives PCs and stroke prevalence had a DIC of 5103 when spatial random effects were included in the model when spatial random effects were not included DIC=5129. However, for the model examining associations between aesthetic objective PCs from private water supplies and hypertension prevalence, the value for pD was -32.8 when the spatial random effects were included, making comparisons between DICs invalid. With the exception of the model examining associations between arsenic in public supplies and hypertension (pD=9.2), all other models for hypertension prevalence had a negative pD (range of -11.2 to -71.4).



**Table 6.25** Values for Moran's I for global spatial autocorrelation for residuals from frequentist models for the association between measures of water quality and cardiovascular disease from public and private water supplies.

Model	Hypertension		Ischemic Heart Disease		Stroke	
	Moran's I	p	Moran's I	p	Moran's I	p
<b><i>Public water supplies</i></b>						
Arsenic	0.223	0.001	0.208	0.001	0.100	0.02
Health Standards PCs	0.164	0.001	0.135	0.007	0.070	0.06
Aesthetic Objective PCs	0.239	0.001	0.115	0.006	0.026	0.26
<b><i>Private wells</i></b>						
Arsenic	0.237	0.001	0.211	0.001	0.101	0.02
Health Standards PCs	0.234	0.001	0.144	0.004	0.095	0.02
Aesthetic Objective PCs	0.242	0.001	0.147	0.001	0.089	0.02
Null model	0.228	0.001	0.222	0.001	0.106	0.008

PC = principal component

Risk factors included arsenic concentrations, health standards PC scores, and aesthetic objective PC scores. Each model was adjusted for sex, age, prevalence of COPD, education level and income. Residuals for null models or models for prevalence with no fixed effects were also analyzed.

#### 6.4.12 Prior sensitivity

Sensitivity analysis for alternative specifications for the hyperpriors on precision for the unstructured and spatially correlated random effects was evaluated for the model examining associations between health standards PCs in public water supplies and stroke prevalence, the sole Bayesian model in which a positive association was identified between a water risk factor and a health outcome. The analysis indicated that the estimates for regression coefficients were very similar regardless of the hyperpriors used for this outcome (Table 6.26).

**Table 6.26** Results of prior sensitivity analysis comparing the reported Bayesian model of associations between health standards principal components in public water supplies and stroke prevalence in 2010 to models with two alternative hyperprior specifications for the precision of the unstructured and structured random effects.

Effect	Reported model			Alternative hyperpriors on random effects					
	$\sim \text{gamma}(0.001, 0.001)$			$\text{Gamma} \sim (0.5, 0.0005)$			$\text{precision} = 1/SD^2$ where $SD \sim \text{uniform}(0, 5)$		
	95% CrI			95% CrI			95% CrI		
	IRR	lower	upper	IRR	lower	upper	IRR	lower	Upper
Intercept	0.918	0.834	1.012	0.919	0.836	1.010	0.918	0.833	1.010
PC1 <sub>health</sub> Quintile 1	Reference category								
PC1 <sub>health</sub> Quintile 2	1.080	0.970	1.203	1.079	0.971	1.201	1.079	0.969	1.204
PC1 <sub>health</sub> Quintile 3	1.092	0.977	1.222	1.092	0.979	1.220	1.093	0.976	1.224
PC1 <sub>health</sub> Quintile 4	1.050	0.934	1.177	1.053	0.940	1.179	1.051	0.934	1.180
PC1 <sub>health</sub> Quintile 5	1.115	0.992	1.253	1.117	0.996	1.254	1.116	0.991	1.254
PC2 <sub>health</sub>	0.917	0.826	1.019	0.916	0.824	1.017	0.916	0.822	1.021
PC3 <sub>health</sub>	1.012	0.919	1.115	1.011	0.920	1.110	1.013	0.918	1.119
COPD prevalence	2.212	1.256	3.900	2.236	1.260	3.931	2.194	1.242	3.869
Education <sup>1</sup>	0.843	0.721	0.987	0.841	0.720	0.982	0.845	0.722	0.988
Income <sup>2</sup>	0.998	0.995	1.001	0.998	0.995	1.001	0.998	0.995	1.001
Random effects	SD		SE	SD		SE	SD		SE
Unstructured RE	0.090		0.029	0.079		0.037	0.086		0.033
Spatially structured RE	0.081		0.032	0.078		0.037	0.087		0.035

Burn in period = 100,000 iterations

SD=standard deviation, CRI = credible interval, PC = principal component, COPD = chronic obstructive pulmonary diseases, SE = standard error, RE = random effect

<sup>1</sup>Proportion of residents who did not complete high school. <sup>2</sup>Average total income (thousands of dollars).

## 6.5 Discussion

This study examined whether the differences between observed and expected age and sex-based risks for cardiovascular disease in geographic regions across the southern regions of the province were associated with measures of water quality in rural Saskatchewan. Associations between arsenic and hypertension and cardiovascular disease outcomes were evaluated as the primary objective. The second and third objectives explored the potential for associations between groups of parameters measured as health standards and groups of parameters measured as aesthetic objectives, as defined in Saskatchewan Drinking Water Quality Standards and Objectives, and hypertension and cardiovascular disease.

No associations were evident between arsenic concentrations in groundwater from public water supplies or private wells and increased risk of any of the outcomes examined in this study. Although some studies have identified a link between arsenic in drinking water and an increased risk of hypertension even at relatively low concentrations of arsenic (Gong and O'Bryant, 2012; Moon et al., 2013; James et al., 2015), the results of the present study are consistent with a recent meta-analysis of the association between arsenic exposure and hypertension which concluded that evidence was inconclusive (Abir et al., 2011) for low concentrations of arsenic. A paradoxical protective effect of arsenic concentrations in public water supplies on hypertension was apparent in the frequentist GLMMs for four of the five years examined. However, this effect was not evident in the Bayesian model for 2010, and the relationship was clearly non-monotonic. This apparent protective association was likely attributable to unmeasured spatially correlated confounding that was accounted for by the inclusion of the spatial random effects in the Bayesian model.

Associations between arsenic and cardiovascular disease have been the subject of recent reviews (Navas-Acien et al., 2005; Wang et al., 2007; Moon et al., 2013; Tsuji et al., 2014). Evidence for a link between drinking water arsenic and cardiovascular disease are not clear, especially at lower concentrations of arsenic, and comparison of studies is hampered by a lack of consistency around measured endpoints, definitions of various types of cardiovascular disease, and control for confounding. In the present study, no evidence was found for associations between arsenic concentrations in groundwater and ischemic heart disease or stroke in residents of rural Saskatchewan.

Arsenic is listed as a health standard in the Saskatchewan Drinking Water Quality Standards and Objectives so was also one of the seven chemicals evaluated with the health standards. Increases in only one of the health standards principal components were associated with an increased risk of disease in the Bayesian models. An increased PC1<sub>health</sub> score in public water supplies was associated with an increase in stroke prevalence in the Bayesian model for 2010 when PC1<sub>health</sub> was evaluated as a continuous variable. Although a strong dose-response relationship was not evident when the variable was categorized as quintiles, this effect could warrant further investigation given the small range of exposures estimated by kriging. The first principal component from public water supplies was characterized primarily by selenium and nitrate

concentrations, with a smaller contribution from lead. It has been suggested lead increases the risk of cardiovascular disease (Lustberg and Silbergeld, 2002; Bhatangar, 2006; Navas-Acien et al., 2007) which supports the findings in the present study. Selenium supplementation has been hypothesized to protect against cardiovascular disease (Oster and Prellwitz, 1990) which is opposite the effect noted for PC1<sub>health</sub> in the present study; however, recent trials suggest there is no benefit to selenium supplementation in the prevention of cardiovascular disease (Rees et al., 2013). However, interpretation of the health standards PCs from public supplies requires caution. The value for the Kaiser-Meyer-Olkin measure of sampling adequacy for health standards in public supplies was just below the minimum value recommended for PCA, suggesting the amount of variance shared by the parameters available for PCA was low (Chapter 4).

In the Bayesian models for 2010, there were no other associations between any of the principal components derived from the health standards for drinking water and increased risks of hypertension or cardiovascular disease. In contrast, several apparent protective associations were identified among frequentist models. Because these effects were not apparent in the Bayesian spatial models, it is likely that unmeasured, spatially correlated confounders accounted for the protective associations apparent in the frequentist models. For example, for public water supplies, PC1<sub>health</sub> was associated with decreased risk of IHD in 4 of 5 years, opposite to its apparent association with increased risk of stroke in the Bayesian model. Also from public supplies, PC2<sub>health</sub> (characterized by positive loading of boron and a strong negative loading of barium) was associated with decreased risk of hypertension. In addition, the previously discussed low measures of sampling adequacy for health standards in public supplies mean these results require cautious interpretation. For private water supplies, PC1<sub>health</sub> (characterized by selenium, nitrate and uranium) was associated with a decreased risk of IHD. With the exceptions of a questionable protective effect of selenium (Oster and Prellwitz, 1990; Rees et al., 2013), which contributed strongly to PC1<sub>health</sub> in both public and private supplies, biologically plausible explanations for protective effects for the parameters contributing to health standards on the outcomes examined in the present study could not be identified in the literature.

For those parameters assessed in the Saskatchewan Drinking Water Quality Standards and Objectives as aesthetic objectives, those associated with the second principal components showed consistent associations with decreased risk of cardiovascular disease, a finding supported

by the existing literature (Rubenowitz et al., 2000; Monarca et al., 2006; Yang et al., 2006; Catling et al., 2008). For both public and private water supplies, PC2<sub>aesthetic</sub> was characterized by strong contributions from hardness and magnesium. In the Bayesian analysis for 2010, PC2<sub>aesthetic</sub> from public water supplies was associated with a decreased risk of IHD and PC2<sub>aesthetic</sub> from private water supplies was associated with a decreased risk of hypertension. The protective association between PC2<sub>aesthetic</sub> and IHD was also consistently mirrored in the frequentist models for public water supplies. The consistency of this result with other studies, particular those examining the protective relationship between magnesium in drinking water and decreased cardiovascular disease (Rubenowitz et al., 2000; Yang et al., 2006) suggests that the present study had the power to detect associations between water quality parameters and disease. Although a strong dose-response relationship was not evident upon categorization of PC2<sub>aesthetic</sub>, atypical dose-response curves are possible (Calabrese et al., 2001, Vandenberg et al., 2012). This could also reflect the combination of a relative small effect estimate combined with a narrow range of exposures estimated in the present study.

In the frequentist models, consistent protective effects were also apparent between PC1<sub>aesthetic</sub> from public and private supplies and IHD. PC1<sub>aesthetic</sub> shared contributions from sodium, chloride, sulfate, alkalinity and total dissolved solids in both public supplies and private wells. The protective effects apparent in the frequentist models were not apparent in the corresponding Bayesian models, and therefore may be attributed to unmeasured spatial confounders, particularly given the lack of support in the literature for plausible mechanisms to explain protective effects between this component and cardiovascular disease.

Increased PC3<sub>aesthetic</sub> scores from public water supplies were associated with decreased stroke prevalence in the Bayesian analysis for 2010 as well as three of five years in the frequentist models. PC3<sub>aesthetic</sub> was characterized by strong loadings of iron and manganese in both public and private supplies, although no associations were evident between PC3<sub>aesthetic</sub> and stroke in private supplies. A plausible mechanism to explain a protective effect of iron and manganese on cardiovascular disease could not be identified in the literature. Nevertheless, the consistency of protective effects identified between aesthetic objective PC scores and cardiovascular disease suggest that further investigation of the health benefits of water with high concentrations of

parameters identified as aesthetic objectives is important to inform recommendations around the use of water treatment to remove minerals from drinking water.

In the weight of evidence approach used in the present study to evaluate associations between water quality and hypertension, IHD and stroke, the hierarchical Bayesian models were considered primary evidence for associations, because of their ability to adjust for spatially correlated confounders. The tests for global spatial autocorrelation in the random effect residuals from frequentist models consistently suggested the presence of unmeasured, spatially correlated effects on each of the disease outcomes analyzed, supporting the use of spatial models to analyze associations between risk factors and these diseases. Comparison of DICs in Bayesian models with and without the spatial random effect also indicated that spatial models provided better fit for these data. However, the consistent finding of a negative value for pD in the spatial Bayesian models for hypertension prevalence suggested that there were issues with the parameterization of these models, necessitating caution in their interpretation. Because choice of hyperpriors can affect pD (Speigelhalter et al., 2002), sensitivity to alternative hyperpriors was checked for the model examining associations between health standards in public water supplies and hypertension. With the alternative hyperpriors, the effect estimates were consistent with those reported in Table 6.7 and pD remained negative (results not shown), suggesting that the choice of hyperprior was not the factor driving the negative values for pD.

While the health standard PCs were not directly comparable between public and private supplies, arsenic was comparable and the aesthetic objective PCs were quite similar in composition, yet few effects were consistent between public and private water supplies. While this could reflect lack of consistency in hypothesized associations, it is also possible that it reflects differences in how consumers of public and private supplies perceive the safety of their tap water (Chapters 2, Chapter 3). Different tap water consumption or treatment behaviors between owners of private water supplies and public water supplies could affect our exposure assessment and confound the observed associations between water quality and hypertension and cardiovascular disease.

In comparison, there was less variability in the results for the covariates included in the models as potential confounders (COPD prevalence, education and income). While there was some variation year-to-year, the results were relatively consistent for these variables between public and private water supplies, and between frequentist and Bayesian models.

The present study shares limitations with other ecological studies, particularly the potential to be affected by the ecological fallacy. Because the water quality risk factors were assessed at an ecological level, we do not know the exposure status of cases compared to non-cases.

Our exposure assessment was also hampered by not knowing whether the study participants consume water from a private well or public water supply, or if residents treat their water in-home or prefer bottled water (Chapter 3). As a result, even residents using the same water supply could have substantially different personal exposure histories based on consumption patterns and in-home treatment choices.

Estimating exposures from drinking water over large areas, is a potential source of misclassification bias, particularly where there is heterogeneity in the composition of water supplies (Chapter 4). The combination of techniques used to summarize large data sets of water monitoring data resulted in modulation of extreme values (Chapter 4), resulting in low variation in the exposure estimates used in our analysis. For example, arsenic concentrations predicted by kriging rarely exceeded the Saskatchewan drinking water standard of 10 µg/L, and after taking the mean predicted value over the geographic units the arsenic concentrations used in the analyses were well below 10 µg/L. A recent systematic review used >50µg/L as the high exposure category in a pooled analysis of relative risks for cardiovascular disease (Moon et al., 2012); the same review concluded that the evidence for links between arsenic and cardiovascular disease was inconclusive at lower concentrations. The potential misclassification of our exposure estimates likely resulted in bias towards the null.

The attribution of residential location data in the administrative database is another potential source of misclassification, particularly among First Nations persons (Chapter 5). Our exposure estimate also depended on the assumption that the place of residence in each year represented historical exposure to the water in that area. Although our study population was relatively non-mobile over the study period (Chapter 5), it is likely that exposures were misclassified for some residents.

Although exposure assessment is subject to misclassification in ecological analyses, the potential relevance of long term environmental exposure or distant past exposures in the development of cardiovascular diseases makes exposure assessment challenging even in individual-level studies. Accurate assessment of historical and cumulative exposure is difficult. This is compounded by a

poor understanding of the relevant induction period, which can represent a form of non-differential classification that results in bias toward the null and underestimation of real effects (Rothman, 1981). For arsenic, it has been suggested that even in utero exposure could play a role in the development of cardiovascular disease later in life (Smith and Steinmaus, 2009; Farzan et al., 2013; Abdul et al., 2015). While assessing exposure at an individual level is ideal for developing causal evidence for associations between environmental exposures and disease, such studies are resource intensive (Elliot and Savitz, 2008). While caution is warranted in the interpretation of ecological epidemiological studies, they can be a valuable tool for hypothesis screening and for determining the need for more detailed confirmatory investigations of associations between environmental exposures and chronic diseases (Rothman et al., 2008). This is especially true for exposures with potentially lengthy latent periods (de Vocht et al., 2013).

Case definition for each of the outcomes is a potential source of misclassification and makes comparison of results of the present study and other studies examining associations between arsenic concentrations in drinking water and hypertension, IHD and stroke difficult. Studies investigating associations between arsenic and hypertension and cardiovascular disease have used a range of case definitions including self-reported, clinical measures of blood pressure, medication use, death certificates, and subclinical markers of disease including electrocardiogram changes and carotid artery intimal-medial thickness (Navas-Acien et al., 2007; Wang et al., 2007; Abir et al., 2011; Moon et al., 2012). Few studies have used administrative data or ICD codes to define cases of illness with the exception of studies using mortality outcomes based on death certificates. While the algorithms used in the present study have been validated, there is potential for misclassification of the disease outcomes using administrative data; the sensitivity of the algorithms indicates that the prevalence of each of the outcomes was likely underestimated. The algorithm used to identify cases of hypertension was reported to have a sensitivity of 75% and specificity of 94%, (Quan et al., 2009), while the algorithm used to identify cases of IHD had a sensitivity of 72.4% and a specificity of 97.6% (Tu et al., 2010) and the algorithm used to identify cases of stroke had a sensitivity of 60.2% and a specificity of 99.2% (Tu et al., 2013). In 2007, chronic disease management codes were introduced, allowing physicians submitting fee claims to include all chronic diseases being treated. Prior to 2007 physician visits were coded according to the primary reason for a visit, potentially leading to underestimation of prevalence of chronic conditions such as hypertension from physician billing



records. By using a minimum run-in period of four years for case identification, and focussing the Bayesian analysis on 2010 data, allowing an eight year run in for identifying cases, the potential for underestimating prevalence in the present study was minimized.

Data on important risk factors for hypertension and cardiovascular disease, including body mass index, physical inactivity, dietary choices, alcohol consumption, psychosocial factors, and serum lipid profiles (Yusuf et al., 2004; Public Health Agency of Canada, 2009; O'Donnell et al., 2010; D'Agostino et al., 2013), were not available for our study population. While data on some of these risk factors are available in the Canadian Community Health Survey (Statistics Canada, 2015), the resolution at which these data were available in the study population was too coarse to add useful information to the present analyses. Data on these risk factors would be important to incorporate into future studies.

We controlled for smoking rates using stratum specific COPD prevalence as a surrogate measure for smoking history. COPD is strongly associated with smoking history (Single et al., 2000) and cases could be summarized from the administrative health data as a covariate measure, whereas smoking rates were available only at a coarse resolution. Smoking is an important risk factor for each of our outcomes, which was expected to be reflected in the model results even though not the focus of analysis. COPD prevalence was associated with increased prevalence of IHD in all years, but consistently with stroke prevalence only in the 2009-2010 models. COPD was not associated with hypertension, which was unexpected. The consistency of expected associations between COPD and IHD suggests that the model interpretation was most supported for the IHD models, less supported for stroke, and reinforces the need for caution in the interpretation of the hypertension results.

Socioeconomic variables were included as covariates at the ecological level; however, in rural areas these data are subject to measurement error due to methods Statistics Canada uses to protect confidentiality in small populations (Chapter 5). Although socioeconomic status is considered an important risk factor with an inverse relationship to cardiovascular disease (Kaplan and Keil, 1993), associations between the socioeconomic variables and the outcomes were not significant in many of the models in the present study. Where significant associations were demonstrated, higher income was associated with lower prevalence of the outcomes as expected, but lower educational attainment was often associated with lower prevalence, opposite

of what was expected. It is possible that the completion of high school is not a sensitive measure of the impact of education on the prevalence of the outcomes in the study population. These results could also be related to measurement error due to suppression strategies used for publically available data.

Diabetes and hypertension are also important risk factors for cardiovascular outcomes including ischemic heart disease and stroke (Yusuf et al., 2004; Public Health Agency of Canada, 2009; O'Donnell et al., 2010; D'Agostino et al., 2013), but were not included as risk factors in the models. Bayesian modeling of hypertension and diabetes as joint outcomes with IHD or stroke in future analyses is recommended as a means of better understanding the role of water quality as risk factors for these diseases.

This study used previously collected administrative health data and water quality surveillance data to analyze associations between arsenic and other aspects of water quality and hypertension and cardiovascular disease. Although there were limitations to the data quality, individual level data was incorporated by stratification and indirect standardization for sex and age, improving the study over a strictly ecological approach (Kunzli and Tager, 1997; Elliot and Savitz, 2008). A long run-in period for each of the case definitions was employed to minimize the chances of underestimating prevalence. Hierarchical Bayesian models incorporating independent and spatially correlated random effect were used as the primary evidence for associations, owing to their strength at accounting for unmeasured but spatially correlated confounders.

## **6.6 Conclusions**

Based on the present study, there does not appear to be an indication that groundwater arsenic concentrations are associated with population-level increased risk of hypertension or cardiovascular disease among rural Saskatchewan residents. Further investigation of an association identified between a principal component associated most strongly with selenium, nitrate and lead in public supplies and an increased risk of stroke may be warranted. It is possible that small effects of arsenic or other aspects of water quality on increased risks of hypertension and cardiovascular disease were not detected due to the ecological nature of the present study. The most striking finding was an apparent association between the principal component represented primarily by water hardness and magnesium, a decreased risk of IHD and hypertension. The consistency of these finding across multiple models in the present study and

with other studies warrants further investigation owing to possible implications for public health. It is possible that consumption of water with elevated concentrations of some minerals has a protective effect on cardiovascular disease; treatment of household drinking water with methods to remove minerals could increase the risk of cardiovascular disease. To better evaluate the effects of water quality on hypertension and cardiovascular disease, exposure and outcome assessment at an individual level is recommended. Exposure assessments should take individual consumption patterns and in-home treatment methods into account.

## 6.7 References

- Abdul, K.S.M., Jayasinghe, S.S., Chandana, E.P.S., Jayasumana, C., De Silva, P.M.C.S., 2015. Arsenic and human health effects: A review. *Environmental Toxicology and Pharmacology* 40, 828–846. doi:10.1016/j.etap.2015.09.016
- Abhyankar, L.N., Jones, M.R., Guallar, E., Navas-Acien, A., 2012. Arsenic Exposure and Hypertension: A Systematic Review. *Environmental Health Perspectives* 120, 494–500.
- Abir, T., Rahman, B., D’Este, C., A, F., Milton, A.H., 2012. The Association between Chronic Arsenic Exposure and Hypertension: A Meta-Analysis. *Journal of Toxicology* 2012, Article ID 198793. doi:10.1155/2012/198793
- Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1–20. doi:10.1007/BF00116466
- Bhatnagar, A., 2006. Environmental Cardiology Studying Mechanistic Links Between Pollution and Heart Disease. *Circ Res* 99, 692–705. doi:10.1161/01.RES.0000243586.99701.cf
- Calabrese, E.J., Baldwin, and L.A., 2001. U-Shaped Dose-Responses in Biology, Toxicology, and Public Health. *Annual Review of Public Health* 22, 15–33. doi:10.1146/annurev.publhealth.22.1.15
- Catling, L.A., Abubakar, I., Lake, I.R., Swift, L., Hunter, P.R., 2008. A systematic review of analytical observational studies investigating the association between cardiovascular disease and drinking water hardness. *Journal of Water and Health* 6, 433–442. doi:10.2166/wh.2008.054
- Chen, C.-J., Chiou, H.-Y., Chiang, M.-H., Lin, L.-J., Tai, T.-Y., 1996. Dose-Response Relationship Between Ischemic Heart Disease Mortality and Long-term Arsenic Exposure. *Arterioscler Thromb Vasc Biol* 16, 504–510. doi:10.1161/01.ATV.16.4.504
- Chen, C.-J., Hsueh, Y.-M., Lai, M.-S., Shyu, M.-P., Chen, S.-Y., Wu, M.-M., Kuo, T.-L., Tai, T.-Y., 1995. Increased Prevalence of Hypertension and Long-term Arsenic Exposure. *Hypertension* 25, 53–60. doi:10.1161/01.HYP.25.1.53
- D’Agostino, R.B., Pencina, M.J., Massaro, J.M., Coady, S., 2013. Cardiovascular Disease Risk Assessment: Insights from Framingham. *Glob Heart* 8, 11–23. doi:10.1016/j.gheart.2013.01.001
- de Vocht, F., Hannam, K., Buchan, I., 2013. Environmental risk factors for cancers of the brain and nervous system: the use of ecological data to generate hypotheses. *Occup Environ Med* 70, 349–356. doi:10.1136/oemed-2012-100954
- Elliott, P., Savitz, D.A., 2008. Design Issues in Small-Area Studies of Environment and Health. *Environ Health Perspect* 116, 1098–1104. doi:10.1289/ehp.10817
- Environment and Climate Change Canada, 2007. Environment and Climate Change Canada - Water - Groundwater. URL <https://www.ec.gc.ca/eau-water/default.asp?lang=En&n=300688DC-1#sub5>.

- Farzan, S.F., Karagas, M.R., Chen, Y., 2013. In utero and early life arsenic exposure in relation to long-term health and disease. *Toxicology and Applied Pharmacology* 272, 384–390. doi:10.1016/j.taap.2013.06.030
- Flora, S.J.S. (Ed.), 2015. *Handbook of arsenic toxicology*. Academic Press, an imprint of Elsevier, Amsterdam.
- Frisoli, T.M., Schmieder, R.E., Grodzicki, T., Messerli, F.H., 2012. Salt and Hypertension: Is Salt Dietary Reduction Worth the Effort? *The American Journal of Medicine* 125, 433–439. doi:10.1016/j.amjmed.2011.10.023
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical science* 457–472.
- Gershon, A.S., Wang, C., Guan, J., Vasilevska-Ristovska, J., Cicutto, L., To, T., 2009. Identifying Individuals with Physician Diagnosed COPD in Health Administrative Databases. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 6, 388–394. doi:10.1080/15412550903140865
- Gong, G., O'Bryant, S.E., 2012. Low-level arsenic exposure, AS3MT gene polymorphism and cardiovascular diseases in rural Texas counties. *Environmental Research* 113, 52–57. doi:10.1016/j.envres.2012.01.003
- Government of Canada, 2014. Canadian Chronic Disease Surveillance System 1996/1997-2011/2012. URL <http://open.canada.ca/data/en/dataset/9525c8c0-554a-461b-a763-f1657acb9c9d>.
- Government of Saskatchewan, 2002. *Drinking Water - Source to Tap Solutions*.
- James, K.A., Byers, T., Hokanson, J.E., Meliker, J.R., Zerbe, G.O., Marshall, J.A., 2015. Association between Lifetime Exposure to Inorganic Arsenic in Drinking Water and Coronary Heart Disease in Colorado Residents. *Environ Health Perspect* 123, 128–134. doi:10.1289/ehp.1307839
- Jang, M.J., Lee, Y., Lawson, A.B., Browne, W.J., 2007. A comparison of the hierarchical likelihood and Bayesian approaches to spatial epidemiological modelling. *Environmetrics* 18, 809–821.
- Kaplan, G.A., Keil, J.E., 1993. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation* 88, 1973–1998.
- Künzli, N., Tager, I.B., 1997. The semi-individual study in air pollution epidemiology: a valid design as compared to ecologic studies. *Environ Health Perspect* 105, 1078–1083.
- Lawson, A.B., 2013. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, Second Edition. Chapman and Hall/CRC, Boca Raton.
- Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique and future directions. *Statist. Med.* 28, 3049–3067. doi:10.1002/sim.3680
- Lustberg, M., Silbergeld, E., 2002. Blood lead levels and mortality. *Arch Intern Med* 162, 2443–2449. doi:10.1001/archinte.162.21.2443

- Malik, V.S., Popkin, B.M., Bray, G.A., Després, J.-P., Hu, F.B., 2010. Sugar-Sweetened Beverages, Obesity, Type 2 Diabetes Mellitus, and Cardiovascular Disease Risk. *Circulation* 121, 1356–1364. doi:10.1161/CIRCULATIONAHA.109.876185
- McLeod, L., Bharadwaj, L., Waldner, C., 2015. Risk factors associated with perceptions of drinking water quality in rural Saskatchewan. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques* 40, 36–46. doi:10.1080/07011784.2014.985513
- Monarca, S., Donato, F., Zerbini, I., Calderon, R.L., Craun, G.F., 2006. Review of epidemiological studies on drinking water hardness and cardiovascular diseases. *European Journal of Cardiovascular Prevention & Rehabilitation* 13, 495–506. doi:10.1097/01.hjr.0000214608.99113.5c
- Moon, K., Guallar, E., Navas-Acien, A., 2012. Arsenic exposure and cardiovascular disease: an updated systematic review. *Current atherosclerosis reports* 14, 542–555.
- Moore, D.F., Lix, L.M., Yogendran, M.S., Martens, P., Tamayo, A., 2008. Stroke surveillance in Manitoba, Canada: Estimates from administrative databases. *Chronic Diseases in Canada* 29, 22–30.
- Navas-Acien, A., Guallar, E., Silbergeld, E.K., Rothenberg, S.J., 2007. Lead Exposure and Cardiovascular Disease—A Systematic Review. *Environ Health Perspect* 115, 472–482. doi:10.1289/ehp.9785
- Navas-Acien, A., Sharrett, A.R., Silbergeld, E.K., Schwartz, B.S., Nachman, K.E., Burke, T.A., Guallar, E., 2005. Arsenic Exposure and Cardiovascular Disease: A Systematic Review of the Epidemiologic Evidence. *Am. J. Epidemiol.* 162, 1037–1049. doi:10.1093/aje/kwi330
- O'Donnell, M.J., Xavier, D., Liu, L., Zhang, H., Chin, S.L., Rao-Melacini, P., Rangarajan, S., Islam, S., Pais, P., McQueen, M.J., others, 2010. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *The Lancet* 376, 112–123.
- Onufrak, S.J., Park, S., Sharkey, J.R., Sherry, B., 2014. The relationship of perceptions of tap water safety with intake of sugar-sweetened beverages and plain water among US adults. *Public Health Nutrition* 17, 179–185. doi:10.1017/S1368980012004600
- Oster, O., Prellwitz, W., 1990. Selenium and cardiovascular disease. *Biol Trace Elem Res* 24, 91–103.
- Padwal, R., Campbell, N., Touyz, R.M., 2005. Applying the 2005 Canadian Hypertension Education Program recommendations: 3. Lifestyle modifications to prevent and treat hypertension. *CMAJ* 173, 749–751. doi:10.1503/cmaj.050186
- Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R news* 6, 7–11.
- Public Health Agency of Canada, 2014. Economic Burden of Illness in Canada, 2005–2008. URL <http://www.phac-aspc.gc.ca/publicat/ebic-femc/2005-2008/assets/pdf/ebic-femc-2005-2008-eng.pdf>.

- Public Health Agency of Canada, 2009. 2009 Tracking Heart Disease and Stroke in Canada - Public Health Agency of Canada.
- Quan, H., Chen, G., Tu, K., Bartlett, G., Butt, D.A., Campbell, N.R.C., Hemmelgarn, B.R., Hill, M.D., Johansen, H., Khan, N., Lix, L.M., Smith, M., Svenson, L., Walker, R.L., Wielgosz, A., McAlister, F.A., 2013. Outcomes Among 3.5 Million Newly Diagnosed Hypertensive Canadians. *Canadian Journal of Cardiology* 29, 592–597. doi:10.1016/j.cjca.2012.12.016
- Quan, H., Khan, N., Hemmelgarn, B.R., Tu, K., Chen, G., Campbell, N., Hill, M.D., Ghali, W.A., McAlister, F.A., 2009. Validation of a Case Definition to Define Hypertension Using Administrative Data. *Hypertension* 54, 1423–1428. doi:10.1161/HYPERTENSIONAHA.109.139279
- Rahman, M., Tondel, M., Ahmad, S.A., Chowdhury, I.A., Faruquee, M.H., Axelson, O., 1999. Hypertension and arsenic exposure in Bangladesh. *Hypertension* 33, 74–78.
- Rees, K., Hartley, L., Day, C., Flowers, N., Clarke, A., Stranges, S., 2013. Selenium supplementation for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 31, CD009671. doi:10.1002/14651858.CD009671.pub2
- Robitaille, C., Bancej, C., Dai, S., Tu, K., Rasali, D., Blais, C., Plante, C., Smith, M., Svenson, L.W., Reimer, K., Casey, J., Puchtinger, R., Johansen, H., Gurevich, Y., Waters, C., Lix, L.M., Quan, H., 2013. Surveillance of ischemic heart disease should include physician billing claims: population-based evidence from administrative health data across seven Canadian provinces. *BMC Cardiovascular Disorders* 13, 88. doi:10.1186/1471-2261-13-88
- Robitaille, C., Dai, S., Waters, C., Loukine, L., Bancej, C., Quach, S., Ellison, J., Campbell, N., Tu, K., Reimer, K., Walker, R., Smith, M., Blais, C., Quan, H., 2012. Diagnosed hypertension in Canada: incidence, prevalence and associated mortality. *CMAJ* 184, E49–E56. doi:10.1503/cmaj.101863
- Rothman, K.J., 1981. Induction and Latent Periods. *Am. J. Epidemiol.* 114, 253–259.
- Rothman, K.J., Greenland, S., Lash, T.L., 2008. *Modern Epidemiology*. Lippincott Williams & Wilkins, Philadelphia, PA.
- Rubenowitz, E., Molin, I., Axelsson, G., Rylander, R., 2000. Magnesium in drinking water in relation to morbidity and mortality from acute myocardial infarction. *Epidemiology* 11, 416–421.
- Single, E., Rehm, J., Robson, L., Truong, M.V., 2000. The relative risks and etiologic fractions of different causes of death and disease attributable to alcohol, tobacco and illicit drug use in Canada. *CMAJ* 162, 1669–1675.
- Smith, A.H., Steinmaus, C.M., 2009. Health Effects of Arsenic and Chromium in Drinking Water: Recent Human Findings. *Annu Rev Public Health* 30, 107–122. doi:10.1146/annurev.publhealth.031308.100143

- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 583–639.
- Statistics Canada, 2015a. Canadian Community Health Survey - Annual Component (CCHS). URL <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226>.
- Statistics Canada, 2015b. Leading causes of death, by sex (Both sexes). URL <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/hlth36a-eng.htm>.
- Statistics Canada, 2014. 2006 Census Topic-based tabulations. URL <http://www12.statcan.gc.ca/census-recensement/2006/dp-pd/tbt/index-eng.cfm>.
- Statistics Canada, 2008a. 2006 Census: Data quality and confidentiality standards and guidelines (public): Data quality practices. URL <https://www12.statcan.gc.ca/census-recensement/2006/ref/notes/DQ-QD/DQPractices-QDPratiques-eng.cfm>.
- Statistics Canada, 2008b. Data quality and confidentiality standards and guidelines (public): Confidentiality (non-disclosure) rules. URL <https://www12.statcan.gc.ca/census-recensement/2006/ref/notes/DQ-QD/confidentiality-confidentialite-eng.cfm>.
- Thompson, T.S., 2003. General Chemical Water Quality of Private Groundwater Supplies in Saskatchewan, Canada. *Bulletin of Environmental Contamination and Toxicology* 70, 0447–0454. doi:10.1007/s00128-003-0007-3
- Thompson, T.S., Le, M.D., Kasick, A.R., Macaulay, T.J., 1999. Arsenic in Well Water Supplies in Saskatchewan. *Bulletin of Environmental Contamination and Toxicology* 63, 478–483. doi:10.1007/s001289901005
- Tseng, C.-H., Chong, C.-K., Tseng, C.-P., Hsueh, Y.-M., Chiou, H.-Y., Tseng, C.-C., Chen, C.-J., 2003. Long-term arsenic exposure and ischemic heart disease in arseniasis-hyperendemic villages in Taiwan. *Toxicology Letters* 137, 15–21. doi:10.1016/S0378-4274(02)00377-6
- Tsuji, J.S., Perez, V., Garry, M.R., Alexander, D.D., 2014. Association of low-level arsenic exposure in drinking water with cardiovascular disease: A systematic review and risk assessment. *Toxicology* 323, 78–94. doi:10.1016/j.tox.2014.06.008
- Tu, K., Campbell, N.R., Chen, Z.-L., Cauch-Dudek, K.J., McAlister, F.A., 2007. Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 1, e18–e26.
- Tu, K., Mitiku, T., Lee, D.S., Guo, H., Tu, J.V., 2010. Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD). *Canadian Journal of Cardiology* 26, e225–e228. doi:10.1016/S0828-282X(10)70412-8



- Tu, K., Wang, M., Young, J., Green, D., Ivers, N.M., Butt, D., Jaakkimainen, L., Kapral, M.K., 2013. Validity of Administrative Data for Identifying Patients Who Have Had a Stroke or Transient Ischemic Attack Using EMRALD as a Reference Standard. *Canadian Journal of Cardiology* 29, 1388–1394. doi:10.1016/j.cjca.2013.07.676
- Tuthill, R.W., Calabrese, E.J., 1979. Elevated Sodium Levels in the Public Drinking Water as a Contributor to Elevated Blood Pressure Levels in the Community. *Archives of Environmental Health* 34, 197–203.
- Vandenberg, L.N., Colborn, T., Hayes, T.B., Heindel, J.J., Jacobs Jr, D.R., Lee, D.-H., Shioda, T., Soto, A.M., vom Saal, F.S., Welshons, W.V., others, 2012. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocrine reviews* 33, 378–455.
- Wang, C.H., Hsiao, C.K., Chen, C.L., Hsu, L.I., Chiou, H.Y., Chen, S.Y., Hsueh, Y.M., Wu, M.M., Chen, C.J., 2007. A review of the epidemiologic literature on the role of environmental arsenic exposure and cardiovascular diseases. *Toxicology and applied pharmacology* 222, 315–326.
- Wang, S.-L., Li, W.-F., Chen, C.-J., Huang, Y.-L., Chen, J.-W., Chang, K.-H., Tsai, L.-Y., Chou, K.-M., 2011. Hypertension incidence after tap-water implementation: A 13-year follow-up study in the arseniasis-endemic area of southwestern Taiwan. *Science of The Total Environment* 409, 4528–4535. doi:10.1016/j.scitotenv.2011.07.058
- Water Security Agency, n.d. Saskatchewan's Drinking Water Quality Standards and Objectives (Summarized). URL <http://www.saskh2o.ca/pdf/epb507.pdf>
- World Health Organization, 2013. A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis.
- World Health Organization, I.S. of H.W.G., 2003. 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. *Journal of hypertension* 21, 1983–1992.
- Yang, C.-Y., Chang, C.-C., Tsai, S.-S., Chiu, H.-F., 2006. Calcium and magnesium in drinking water and risk of death from acute myocardial infarction in Taiwan. *Environmental Research* 101, 407–411. doi:10.1016/j.envres.2005.12.019
- Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J., others, 2004. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *The Lancet* 364, 937–952.
- Yusuf, S., Reddy, S., Ôunpuu, S., Anand, S., 2001. Global Burden of Cardiovascular Diseases. *Circulation* 104, 2746–2753. doi:10.1161/hc4601.099487

## **CHAPTER 7: CONCLUSIONS**

Household water supplies in rural Saskatchewan are not tested for quality and safety with the same intensity as water supplies in urban centers. Residents who use private water supplies bear sole responsibility for ensuring the quality and safety of their supplies, but may lack the resources to ensure optimal water quality especially if they must rely on source water that has characteristics making it difficult to treat. Similarly, smaller public water supplies may have more restricted options for source water and face challenges accessing the technology and personnel required to bring the water quality to within the standards and objectives established by the province of Saskatchewan for regulated water supplies. The disparity in testing and potentially the quality of water available to residents of rural and remote areas of Saskatchewan compared to residents of urban areas is a potential public health inequity.

The research outlined in this thesis sought to address questions about water quality in rural Saskatchewan and its impacts on the health of residents of rural Saskatchewan. The perceptions of water quality held by rural Saskatchewan residents, and how these perceptions influence the consumption of tap water and other beverages had not been previously quantified. To obtain a baseline understanding of water quality perceptions as well as choices about drinking water, a questionnaire was administered to a large sample of rural Saskatchewan residents. This was followed by an analysis of groundwater quality in rural areas of Saskatchewan, using existing water quality surveillance data based on the Saskatchewan Drinking Water Quality Standards and Objectives. Principal components analysis was used to summarize groups of contaminants defined as health standards and groups of contaminants defined as aesthetic objectives.

Geostatistical techniques were used to interpolate groundwater quality at unmeasured locations for arsenic concentrations and principal component scores. While this analysis enhanced understanding of patterns of groundwater quality in Saskatchewan, it also served as a method of estimating exposures for an epidemiological study of associations between groundwater quality and chronic disease in rural Saskatchewan. Arsenic concentrations, principal component scores summarizing health standards, and principal component scores summarizing aesthetic objectives were investigated for associations with diabetes, hypertension, ischemic heart disease and stroke.

Previously, little was known about the perceptions held by rural Saskatchewan residents about the quality and safety of their household tap water or their primary choices for drinking water. Responses to a questionnaire distributed to 7500 rural households were analyzed to obtain a

baseline understanding of perceptions of water quality and risk among rural Saskatchewan residents. Of the respondents, 25% reported having at least one aesthetic complaint about their tap water, while 12% believed their tap water was not safe to drink, and 3% believed someone had become ill from consuming their household's tap water. In addition, 36% of respondents were worried that their household water supply would become contaminated in the future. Separate models were developed for each of these concerns, and while different sets of risk factors were identified for each model, there were similarities among them. Having experienced a water advisory such as a boil water order, was a significant predictor in all of the models, lending evidence that personal experience may impact trust in a water source. Having at least one aesthetic complaint about the tap water increased the likelihood of respondents perceiving health risks from their tap water. Increasing age as well as long-term residence in the home reduced the likelihood of having aesthetic complaints or fear of water source contamination, but these factors were not important in perception of health risk. A better understanding of the complex interactions of risk factors affecting perception, and how they change in different contexts, will be important to implementing programs aimed at improving water management and safety in rural areas.

The factors that influenced drinking water consumption choices had similarly complex relationships. Among survey respondents, 31% reported drinking primarily bottled water, while just 61% reported drinking tap water daily. Separate models were developed for each of these drinking water choices. The relationship between having aesthetic concerns and believing the tap water was not safe to drink demonstrated the importance of both aesthetic quality and risk perception in predicting consumption patterns. Those who felt their water was not safe were more likely to choose primarily bottled water and less likely to drink tap water on a regular basis. For these respondents, little difference in the probability of choosing bottled or tap water was attributable to aesthetic qualities. However, for those who felt their tap water was safe, poor aesthetic qualities of the water greatly magnified differences in the likelihood of choosing bottled or tap water. Experiencing drinking water advisories and living in an area for less than 10 years increased the likelihood of primarily consuming bottled water, and decreased the likelihood of routinely consuming tap water. In-home treatment of the tap water was associated with a decreased likelihood of primarily consuming bottled water, but the effect was larger for those using a public water supply. In-home water treatment also increased the likelihood of routine

consumption of tap water, and this effect was greater in magnitude for respondents with aesthetic complaints.

In addition, 48% of respondents reported treating their household's tap water in some way. The factors affecting the choice to treat the tap water in home was less complicated; those who used a private water supply were more likely to treat the tap water, as were those who had children in the home. Believing the water was unsafe to drink reduced the likelihood of treating the water, suggesting that risk perception is not an important motivation for treating water in home, perhaps due to reluctance to consume the water.

Similar to the perceptions of risk, understanding the factors that influence drinking water consumption choices can be important in public health messaging about water management and safety. Understanding the factors that motivate choices about water consumption are also important for risk assessments and exposure estimates in studies of health risks attributable to drinking water. In particular, differences in perceptions of quality and risk along with concomitant differences in consumption patterns of tap water between residents using public and private water supplies could impact results of investigations between water quality and disease.

Water quality surveillance data represents an existing source of data, which can be analyzed to summarize trends in water quality over large geographic areas using multivariate statistics and geostatistical methods. The performance of ordinary, universal, and empirical Bayesian kriging for interpolation of arsenic concentrations and principal component scores for health standards and aesthetic objectives in groundwater were compared for public and private water supply data. Over the large area used in this study, agreement between measured and predicted values were generally low. This appeared to be a consequence of underestimation of high values and overestimation of low values. However, trends in overall water quality and arsenic concentrations were still apparent. Data summarized in this way can inform water quality assessment and management activities and can also inform exposure assessments for investigations of associations between water quality and health outcomes.

The results of the geostatistical analyses from chapter 4 were used in ecological analyses of investigation of associations between water quality and diabetes and cardiovascular disease in

rural residents of southern Saskatchewan. Associations were evaluated between arsenic concentrations and principal component scores for health standards and aesthetic objectives and health outcomes derived from administrative health data, using methods to account for spatial correlation between areas.

No associations were demonstrated between increased arsenic concentrations in groundwater supplies and increased incidence or prevalence of diabetes in rural Saskatchewan. The first principal component score for health standards in public supplies (characterized mainly by selenium, nitrate, and lead) was associated with an increase in diabetes prevalence in 2010 in the Bayesian analysis. However, a lack of support for this association across other time periods as well as the published literature suggested this might have been a spurious finding. An association between increased scores for the first principal component (characterized mainly by sodium, chloride, sulphate, alkalinity, and total dissolved solids) and a decrease in diabetes incidence was limited to the Bayesian model for 2010-2012. Given the lack of any plausible explanation for an association, this finding may also have been spurious. The ecological analysis produced no compelling evidence for any association between problems with groundwater quality and increasing diabetes risk in rural Saskatchewan residents.

Similarly, no associations were evident between elevated arsenic concentrations in groundwater and an increased prevalence of cardiovascular disease in rural Saskatchewan residents. An increase in scores for the first principal component for public supplies was associated with increased prevalence of stroke; this was the only association between a water quality variable and increased risk of a CVD outcome. Increased scores for the second aesthetic objective PC in public supplies were associated with a decreased prevalence of ischemic heart disease. This protective effect was consistently demonstrated across models for different years as well as the Bayesian model. A similar association was demonstrated for the second aesthetic objective PC scores for private wells and decreased prevalence of hypertension. This PC was characterized primarily by hardness and magnesium for both types of water supplies, and the effect demonstrated was plausible given consistency with published literature. An association between the third aesthetic objective PC scores in public supplies (primarily characterized by iron and manganese) and decreased prevalence of stroke was evident in the Bayesian model. This

association was not consistent in frequentist models for all years, and was not supported by the literature.

The results of the ecological investigations of associations between groundwater quality and diabetes and cardiovascular disease were complex. Overall, the concentrations of contaminants defined as health standards in the Saskatchewan Drinking Water Quality Standards and Objectives, including arsenic, do not appear to be associated with the risk of diabetes and cardiovascular disease at the population level. The most compelling result from the current study was the association between increased scores for the second aesthetic objective principal component and a reduced risk of cardiovascular disease. The frequency of values above recommended objective for hardness and magnesium in rural Saskatchewan groundwater combined with the burden of disease from cardiovascular disease warrants further research into this link. While the data necessary to investigate this hypothesis was not available in the current study, treatment to mitigate the presence of minerals in groundwater supplies could inadvertently increase risk of cardiovascular disease in rural populations.

## **7.1 Strengths of the Research**

The robust statistical methods used in this thesis represent unique approaches to addressing the research questions outlined in the objectives. A novel approach, Canada Post's Unaddressed Admail service, was used to distribute questionnaires to a large sample of residents in several large regions of Saskatchewan. The mixed models used to analyze the responses accounted for clustering by postal code and region resulting from the sampling strategy. The results provide a baseline for future assessments of attitudes and behavior towards drinking water quality in Saskatchewan.

The use of principal components analysis combined with kriging to summarize water quality data has been used in a limited number of previous studies which characterized trends in water quality over regions, typically to investigate anthropogenic and natural influences on water quality. However, no studies were identified that validated the use of principal component scores predicted by kriging. Similarly, previous studies using principal component scores as exposure variables to assess associations between water quality and disease outcomes were not identified.

This approach is a feasible way to summarize trends in mixtures of common drinking water contaminants.

The use of administrative health data allowed discernment of cases at a population level for multiple outcomes (i.e. diabetes, hypertension, ischemic heart disease, and stroke) along with the geographic distribution of cases. The stratification of the outcome data by sex and age, and by first Nations status for diabetes cases, allowed for control of these important confounders at the individual level in otherwise ecological analyses.

The exposure-outcome analysis for each of the disease outcomes was performed using Bayesian hierarchical modelling techniques. These models incorporated information on neighboring areas to smooth extreme values in the case counts of the outcomes due to small counts in the stratified outcome data. The inclusion of spatially correlated random effects partitioned the models' residual variance into random variance and variance attributable to location to better account for unmeasured spatially correlated confounding.

Because the Bayesian models were very computationally intense, Bayesian models incorporating spatially correlated as well as uncorrelated random effects were summarized for a single time period. These were compared to frequentist models, which incorporated only uncorrelated random effects, for multiple time periods. Given that residents who become ill may have moved away from rural areas to live in closer proximity to health care providers, and water quality from public supplies likely improved over time, comparison to previous years was important to mitigate the potential loss of cases that were exposed and diagnosed early in the study period, but who left the study population prior to the final year of analysis. Comparison of models over multiple years also allowed for assessment of the stability of effect estimates in different years.

## **7.2 Limitations**

While the response rate for the questionnaire was robust for an anonymously mailed survey (27.5%), potential for biases related to the response rate may preclude the generalizability of the results to the entire rural population of Saskatchewan. In particular, because the questionnaire addressed concerns about water quality and safety, a differential response rate was possible among those that had concerns compared to those who were more confident about their water supply.



Validation of the kriging methods for predicting principal component scores and arsenic concentrations showed that while kriging may interpolate values that reflect trends in water quality, it was not able to accurately predict the full range of principal component scores or arsenic concentrations due to attenuation of extreme values. Averaging the krigged values over the geographic units defined for the exposure-outcome analyses further attenuated the principal component scores and arsenic concentrations used as risk factors in the analyses. As a result, the range of exposures estimated for use in the exposure-outcome analyses was very narrow compared to the values that were observed in the raw data. This exposure attenuation represents a likely source of exposure misclassification that could have biased associations between water quality and diabetes or cardiovascular disease. Due to the attenuation of high values of the exposure variables, it is most likely the direction of bias would have been towards the null.

The output of kriging analysis in ArcGIS includes a standard error map describing uncertainty in the predicted values. This uncertainty could be propagated into the Bayesian regression models to potentially mitigate some bias introduced by exposure misclassification (Lash et al., 2011). However, increasing the model complexity was not practical in the present analysis due to computer hardware limitations in the secure environment in which the health data was housed, and the resulting time needed to run the described models to convergence.

Accurate exposure assessment was also hampered by basing the exposure on area of residence for the study population for each year. If long term exposure to water contaminants was important for induction of the health outcomes, the assumption that residents were exposed to water in their area for some period of time prior to year of analysis could have resulted in exposure misclassification. This bias could have been mitigated somewhat by including only residents who had lived in an area in all previous years of the study period in the yearly cohorts. However, the residence information itself was also subject to misclassification because it was based primarily on postal codes. In rural areas, residents are likely to pick up mail in central locations, which could potentially have been located outside their actual area of residence. As a result, some residents could have been assigned to a study area in which they did not reside.

Administrative health data is a rich data source for identifying cases of disease, but the use of billing codes to identify cases can potentially result in misclassification of the outcomes. Validated, published algorithms were used to identify cases, but these do not have perfect

sensitivity or specificity and errors in data entry at the billing stage are possible, resulting in some uncertainty in the case ascertainment used in the study. This may have been exacerbated for diabetes. Although the goal was to evaluate associations between water quality and type 2 diabetes, some of the administrative data was based on the International Classification of Diseases 9<sup>th</sup> Edition, which does not differentiate type 1 from type 2 diabetes. In addition, physician visits for multiple problems could lead to missed codes in billing records when the primary reasons for a visit was not a pre-existing chronic disease, especially prior to 2007 when multiple chronic disease codes per visit were implemented. This could have resulted in underestimation of disease outcomes, and could also have resulted in misclassification of incident versus prevalent cases of diabetes.

The results of the exposure-outcome analyses were subject to the ecological fallacy. Data about some important confounders were collected and adjusted for at the individual level using indirect standardization, but water quality exposure was only estimated at the area level. The potential for ecological fallacy was magnified somewhat by the separate analyses of public water supplies and private wells. Cases could have consumed water from public supplies, private wells, or a combination of both. Residents also could have modified their exposure by in-home treatment of tap water, or choosing to consume bottled water instead of tap water. Results of the analysis for public water supplies were often inconsistent with the results for private well data. This could reflect differences in the principal components for the different types of supply, but could also reflect different consumption patterns related to perceptions of tap water quality and safety between users of each type of supply.

### **7.3 Future research**

While a baseline was established for understanding perceptions of water quality and risk factors that affect drinking water choices in rural Saskatchewan, future studies could better characterize drinking water consumption patterns among rural residents. A better understanding of the relative amount of tap water and other beverages that are consumed, whether tap water is treated in-home prior to consumption, what motivates the choice to treat water, and which methods are used to treat tap water could enhance future risk assessments. In addition, geographic trends in perception and drinking water choices could be compared to observed water quality to assess congruence between perceptions and general water quality. Studies of risk perception and

associated water testing behavior among residents who use private water supplies would be especially helpful for informing public health education and messaging about water testing.

In other jurisdictions, complex geostatistical methods to predict arsenic concentrations have been explored. While the methods used here are arguably the most practical and accessible to professionals who lack detailed expertise on the use of geographic information systems, future exposure assessments could benefit from development of methods better able to account for local heterogeneity among groundwater supplies. Application of geostatistical methods across more limited geographic regions could potentially improve exposure assessments based on these methods.

Studies that measure both water quality exposure and diabetes or cardiovascular disease outcomes at an individual level would provide more definitive answers to questions about associations between water quality and these diseases. Such studies could be targeted to areas identified by the geostatistical analysis as trending toward high groundwater arsenic concentrations or high principal component scores. Individual level studies could account for the type of water source used as well as individual consumption patterns and choices that could impact exposure assessment, such as in-home water treatment. In particular, individual level studies to further investigate the association between hardness and magnesium and reduced prevalence of cardiovascular disease are recommended.

Associations between arsenic and cancer have also been identified in epidemiological studies in other locations. Individual-level studies between cancer outcomes and arsenic concentrations would also be necessary to more definitively rule out impacts of arsenic in drinking water on cancer risks in rural Saskatchewan.

Because diabetes and hypertension are risk factors for cardiovascular disease, the outcomes evaluated in this study could be closely related. Joint modeling, ideally using Bayesian hierarchical models that account for uncertainty as well as spatial correlation in the outcomes of interest, offer a powerful tool to further investigate the impacts of water quality on these diseases.

Because exposure misclassification and, to a lesser extent, outcome misclassification was a concern in the exposure-outcome assessment, the application of quantitative bias analysis

techniques could provide insight into impact of possible misclassification bias on the results. Quantitative bias analysis could also be used to investigate the potential impact of confounding.

The use of readily available water surveillance data for health standards limited the analysis to aspects of water quality that have been previously established as potential risks. Recognition is growing that a multitude of environmental chemicals can be present in water supplies, including mixtures of substances that could have synergistic effects. The health impacts of these chemicals is largely unknown, but innovative techniques to estimate exposures to mixtures of chemicals and investigate their potential impacts on health will likely become increasingly important for future research.

## **7.4 Summary**

Despite considerable limitations to the analyses of associations between water quality and type 2 diabetes and cardiovascular disease, the manuscripts included in this thesis provide a baseline for investigation of perceptions of water quality and risk, assessment of methods to summarize water quality surveillance data, and investigation of associations between indicators of regional water quality and chronic disease prevalence.

Valuable information about perceptions of water quality and risk among rural residents was obtained. Quantifying the factors influencing perception helped to identify the complexities inherent in estimating exposure for residents using varied types of water supplies.

Summarizing water quality surveillance data over a large geographic region resulted in considerable attenuation of the variability in the observed data, although trends were still evident. This analysis highlighted the challenges of summarizing data over a large scale when considerable local-scale heterogeneity in groundwater characteristics is present. Analysis over smaller scales may be able to improve the interpolation of water quality data for the prediction of water quality at unmeasured locations.

The ecological analysis of associations between water quality and diabetes and cardiovascular disease used a combination of ecologic and individual data sources to evaluate association using population-based data. Bayesian hierarchical modeling is a powerful tool for analysis of area level data and accounts for unmeasured spatial confounding that may lead to spurious results in analyses that do not account for spatial correlation. The use of multivariate statistics to

summarize mixtures of water contaminants as exposure variables for Bayesian analyses of associations between water quality and diabetes and cardiovascular disease represented a unique approach not previously reported. Overall, there was little evidence for any associations between poor water quality in rural Saskatchewan and increasing occurrence of diabetes or cardiovascular disease. However, the observed association between water hardness and magnesium concentrations in groundwater and a reduced risk of CVD warrants further investigation.

## **7.5 References**

Lash, T.L., Fox, M.P., Fink, A.K., 2011. Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer Science & Business Media.

## APPENDIX

### A. Permission to Include Published Manuscripts

#### *A.1 Permission for Chapter 2*

This is the authors accepted manuscript of an article published as the version of record in Canadian Water Resource Journal 13th January 2015: <http://www.tandfonline.com/http://dx.doi.org/10.1080/07011784.2014.985513>

McLeod, L., Bharadwaj, L., Waldner, C., 2015. Risk factors associated with perceptions of drinking water quality in rural Saskatchewan. Canadian Water Resources Journal / Revue canadienne des ressources hydriques 40, 36–46. doi:10.1080/07011784.2014.985513

From Academic Journals Society Permissions <society.permissions@tandf.co.uk> ☆

Subject RE: use of article in dissertation

To Lianne McLeod ☆

24/11/2016 7:04 AM

Reply Forward Archive Junk Delete More

Our Ref: MD/TCWR/P9181

24<sup>th</sup> November 2016

Dear Lianne McLeod,

Thank you for your correspondence requesting permission to reproduce the following article published in our journal in your printed thesis and to be posted in your university's repository at the University of Saskatchewan.

'Risk factors associated with perceptions of drinking water quality in rural Saskatchewan' By Lianne McLeod et al Canadian Water Resource Journal Vol.40:1 (2015) pp.36-46

We will be pleased to grant permission on the sole condition that you acknowledge the original source of publication and insert a reference to the article on the Journals website:

This is the authors accepted manuscript of an article published as the version of record in Canadian Water Resource Journal 13<sup>th</sup> January 2015 <http://www.tandfonline.com/> <http://dx.doi.org/10.1080/07011784.2014.985513>

Please note that this license does not allow you to post our content on any third party websites or repositories.

Thank you for your interest in our Journal.

Yours sincerely

**Michelle Dickson** – Permissions & Licensing Executive, Journals.  
Routledge, Taylor & Francis Group.  
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN, UK.  
Tel: +44 (0)20 7017 7413  
Fax: +44 (0)20 7017 6336  
Web: [www.tandfonline.com](http://www.tandfonline.com)  
e-mail: [Michelle.Dickson@tandf.co.uk](mailto:Michelle.Dickson@tandf.co.uk)

## A.2 Permission for Chapter 3

McLeod, L., Bharadwaj, L., Waldner, C., 2014. Risk Factors Associated with the Choice to Drink Bottled Water and Tap Water in Rural Saskatchewan. *International Journal of Environmental Research and Public Health* 11, 1626–1646. doi:10.3390/ijerph110201626

This manuscript was published in the *International Journal of Environmental Research and Public Health*. Copyright is held by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

---

From IJERPH Editorial Office ★

Subject **Re: Requesting permission to include article in dissertation**

To Lianne McLeod ★

Cc joyce.zhou@mdpi.com <joyce.zhou@mdpi.com> ☆

Dear Lianne,

As you are the author of the manuscript, you are the copyright holder. Thus, no need of permission from our journal.

Best of lucks with your dissertation.

Kind regards,

Unai

--

Unai Vicario PhD  
co-Managing Editor of IJERPH  
MDPI Barcelona Office  
Avenida Madrid, 95, 1 3  
08028 Barcelona, Spain  
Tel. +34 93 639 7662  
[E-Mail:ijerph@mdpi.com](mailto:ijerph@mdpi.com)  
<http://www.mdpi.com/journal/ijerph>

## B. Rural Water Quality Questionnaire – Basis for Chapters 2 and 3



### Drinking Water in Rural Saskatchewan: How do residents in rural Saskatchewan make choices about their drinking water?

1. Where does the tap water in your home come from? *(please check **all** those that apply)*
  - a. community water treatment plant ☐
  - b. public water station / tap ☐
  - c. private water supply located on your property ☐
  - d. other *(please describe)* \_\_\_\_\_ ☐
  - e. there is no tap water in our home ☐
  - f. I don't know ☐
2. Do you know the source(s) of the community water supply, public water station, and/or private water supply to your home?  
YES or NO *(please circle one)*  
If yes, what is the source of your water supply *(please check all that apply)*
  - a. surface water: Lake ☐ River ☐ Creek ☐ Spring ☐  
Dugout, pond or slough ☐ Collected rain water ☐
  - b. water from a underground source: Artesian well ☐ (depth in meters \_\_\_\_\_ or feet \_\_\_\_\_)  
Well with pump ☐ (depth in meters \_\_\_\_\_ or feet \_\_\_\_\_)
  - c. other ☐ *(please describe)* \_\_\_\_\_
3. How does the tap water get to your home? *(please check **all** those that apply)*
  - a. piped into my home from a community water supply ☐
  - b. pumped from a well or surface water source ☐
  - c. piped from artesian well ☐
  - d. delivered by truck or trailer with a water tank ☐
  - e. we have a system to collect rain water ☐
4. Do you have one or more cisterns or tanks at your household for storing drinking water?  
YES or NO *(please circle one)*  
If you have a cistern or tank for drinking water, is it:  
INDOORS and/or OUTDOORS *(please circle all that apply)*
5. Do you ever drink the tap water from your home? *(please check **one**)*
  - a. yes, at least some of the time ☐
  - b. no, we only drink water from other sources (purchased bottles or filled jugs) ☐
6. Do you have any equipment in your home to make the tap water better or safer to drink?  
YES or NO *(please circle one)*  
**If you circled NO, please skip to question 7.**  
If you have equipment is it:
  - a. attached directly to the water pipes: *(check **all** those that apply)*
    - i. Reverse osmosis unit ☐
    - iii. Distilled water system ☐
    - ii. Ultraviolet light system ☐
    - iv. Water softener ☐
  - b. a device/filter under the sink that hooks up to the taps ☐
  - c. water jugs with a filter (Brita, etc.) ☐
  - d. other, please explain \_\_\_\_\_ ☐  
How often do you use the equipment to filter your tap water? *(please circle one)*  
ALWAYS or MOST OF THE TIME or RARELY or NEVER
7. Do you always boil the tap water from your house before you use it for drinking?  
YES or NO *(please circle one)*



8. Why do you filter, boil, or treat tap water in some way before drinking it? *(please check **all** that apply)*
- |   |                          |  |                          |
|---|--------------------------|--|--------------------------|
| a. we do not treat our tap water            | <input type="checkbox"/> | f. filtered/treated water is healthier | <input type="checkbox"/> |
| b. to remove impurities                     | <input type="checkbox"/> | g. prefer filtered water               | <input type="checkbox"/> |
| c. to improve taste                         | <input type="checkbox"/> | h. habit/got used to it                | <input type="checkbox"/> |
| d. to remove chemicals (chlorine, fluorine) | <input type="checkbox"/> | i. was told to do it by someone        | <input type="checkbox"/> |
| e. to ensure safety                         | <input type="checkbox"/> | j. don't know                          | <input type="checkbox"/> |
9. Has your family ever been told to boil the water from the taps in this household before you drink it?  
(Has a boil water order or advisory ever been issued for your current household?)  
YES or NO *(please circle one)*
10. Do you have to boil the water from your taps before you drink it at the time of this questionnaire? (Do you have a boil water order or advisory in place right now?)  
YES or NO *(please circle one)*
- If yes, how long have you had to boil the water? *(please check **one**)*
- |                      |                          |
|----------------------|--------------------------|
| a. less than a week  | <input type="checkbox"/> |
| b. more than a week  | <input type="checkbox"/> |
| c. more than a month | <input type="checkbox"/> |
- If you have been told to boil your water, was this the first time?  
YES or NO *(please circle one)*
- How many **other** times have you been told to boil your water? \_\_\_\_\_
11. Have you ever had a 'do not consume' or a 'do not use' advisory for the tap water in this household?  
YES or NO *(please circle one)*
12. Do you currently have a 'do not consume' or a 'do not use' advisory for the tap water in your household?  
YES or NO *(please circle one)*
- If yes, how long has the advisory been in place? *(please check **one**)*
- |                      |                          |
|----------------------|--------------------------|
| a. less than a week  | <input type="checkbox"/> |
| b. more than a week  | <input type="checkbox"/> |
| c. more than a month | <input type="checkbox"/> |
- Is this the first time these advisories have been in place?  
YES or NO *(please circle one)*
- How many **other** times were you told not to consume or use your tap water? \_\_\_\_\_
13. Do you like drinking the tap water in your home? *(please check **one**)*
- |                                      |                          |
|--------------------------------------|--------------------------|
| a. yes                               | <input type="checkbox"/> |
| b. no                                | <input type="checkbox"/> |
| c. don't care/mind                   | <input type="checkbox"/> |
| d. I do not drink water from the tap | <input type="checkbox"/> |
14. Please check **all** the statements that apply to the tap water from your home:
- |  |                          |
|--|--------------------------|
| a. there is no problem with the tap water              | <input type="checkbox"/> |
| b. I don't think it is safe                            | <input type="checkbox"/> |
| c. the water smells                                    | <input type="checkbox"/> |
| d. the water has a bad taste                           | <input type="checkbox"/> |
| e. it is colored (example it is red, yellow, or brown) | <input type="checkbox"/> |
| f. it is cloudy (not clear)                            | <input type="checkbox"/> |
| g. other comments about my tap water _____             |                          |
15. Do you drink purchased bottled water in your home? *(please check **one**)*
- |   |                          |
|---|--------------------------|
| a. yes, it is the primary drinking water source | <input type="checkbox"/> |
| b. yes, we drink it sometimes                   | <input type="checkbox"/> |
| c. no   | <input type="checkbox"/> |

16. If you drink more tap water than purchased bottled water when you are at home, why?

(please check all that apply)

- a. not applicable (I drink more purchased bottled water) ☐
- b. no difference/ tap water is just as good as bottled water ☐
- c. tap water is more convenient ☐
- d. tap water tastes better ☐
- e. tap water is safer ☐
- f. high cost of bottled water (tap water is cheaper) ☐
- g. I don't like to drink out of plastic bottles ☐
- h. water bottles create more garbage/environmental waste ☐
- i. I don't know why I choose tap water ☐

17. If you drink more purchased bottled water than tap water when you are at home, why?

(please check all that apply)

- a. not applicable (I drink more tap water) ☐
- b. no difference/bottled water is just as good as tap water ☐
- c. bottled water tastes better ☐
- d. bottled water is easier (ready to drink) ☐
- e. bottled water is safer and better for my health ☐
- f. I was told not to drink the tap water ☐
- g. tap water might have bacteria (germs) in it ☐
- h. tap water might have fertilizer and pesticides in it ☐
- i. tap water might have other contaminants in it ☐
- j. my tap water is dirty/cloudy ☐
- k. my tap water has an odor ☐
- l. tap water not available ☐
- m. I don't know why I choose bottled water ☐

18. If no one from your household purchases bottled water, **please check here** ☐, and skip to question 19.

a. The size(s) of bottled water you buy for yourself or your family: (please check all that apply)

- i. less than 1L bottle (example, 500 mL bottle) ☐
- ii. 1L – 2L bottle ☐
- iii. large bottles or jugs (example, for water coolers) ☐
- iv. other: \_\_\_\_\_ ☐

b. If you purchase water, how much money do you spend in a month (on average)? (please check one)

- i. under \$50 ☐
- ii. \$50 - \$100 ☐
- iii. \$100 - \$150 ☐
- iv. more than \$150 ☐
- v. don't know ☐

19. Which of the following beverages do you drink most days at home? (please check all that apply)

- a. tap water ☐
- b. bottled water ☐
- c. pop/soft drink ☐
- d. juice or ice tea (ready to drink) ☐
- e. juice or ice tea (from frozen concentrate) ☐
- f. powder drink mixes ☐
- g. energy drinks (Redbull, Monster) ☐
- h. sports drinks (Gatorade, etc) ☐
- i. milk/chocolate milk ☐
- j. coffee ☐
- k. tea ☐
- l. other: \_\_\_\_\_ ☐

20. Do you worry that you will run out of water at your house?

YES or NO (please **circle one**)

If yes, does this affect your household's water consumption (drinking, cooking, and washing)?

YES or NO (please **circle one**)

21. Do you worry that your drinking water will become contaminated?

YES or NO (please circle one)

22. Are you aware of anything that might currently be affecting the quality your water supply (e.g., pesticides, heavy metals, bacteria, other industrial or agricultural chemicals or contaminants)?

YES or NO (please circle one)

If yes, please explain: \_\_\_\_\_

23. Do you believe that the quality of your drinking water has changed in the past 10 years? Or if you have not lived at your current residence for 10 years, the time that you have been there.

(please check **only one** answer)

- |                                   |                          |
|-----------------------------------|--------------------------|
| a. I don't know                   | <input type="checkbox"/> |
| b. The water quality has improved | <input type="checkbox"/> |
| c. The water quality is the same  | <input type="checkbox"/> |
| d. The water quality has declined | <input type="checkbox"/> |

24. If you **don't** use a community water supply, please check here ☐ , and skip to question 25.

If you use a **community water supply**, do you believe there is adequate and timely testing of the water supply?

YES or NO (please circle one)

25. If you **don't** use a private water supply, please check here ☐ , and skip to question 26.

If you have a **private water supply**, do you test your water supply?

YES or NO (please circle one)

If YES, when is the last time the water was tested: Year: \_\_\_\_\_ Month: \_\_\_\_\_  
(yyyy) (mm)

26. Do you believe there is adequate testing of purchasable bottled water?

YES or NO (please circle one)

27. Have you been sick with vomiting and/or diarrhea in the last 4 weeks?

YES or NO (please circle one)

28. Has any other member of your household been sick with vomiting and/or diarrhea in the last 4 weeks?

YES or NO (please circle one)

If yes, what are the ages of all those who have been sick? \_\_\_\_\_

29. Do you think anyone has ever become sick from drinking the water from your house?

YES or NO or DON'T KNOW (please circle one)

Demographics (This information will help us analyze the results of the survey).

30. Are you?

Male ☐ Female ☐

31. What is your age?

18 – 24	<input type="checkbox"/>	45 – 54	<input type="checkbox"/>
25 – 34	<input type="checkbox"/>	55 – 64	<input type="checkbox"/>
35 – 44	<input type="checkbox"/>	65 or older	<input type="checkbox"/>

32. How many years have you resided in your community?

0 to 5 ☐ 6 to 10 ☐ 11 to 20 ☐ 21 or more ☐

33. Is your household located in a town?

- |   |                          |
|---|--------------------------|
| a. yes                                  | <input type="checkbox"/> |
| b. no, it is an acreage, farm, or ranch | <input type="checkbox"/> |
| c. other, please describe _____         | <input type="checkbox"/> |

34. Total number of people in your household: \_\_\_\_\_

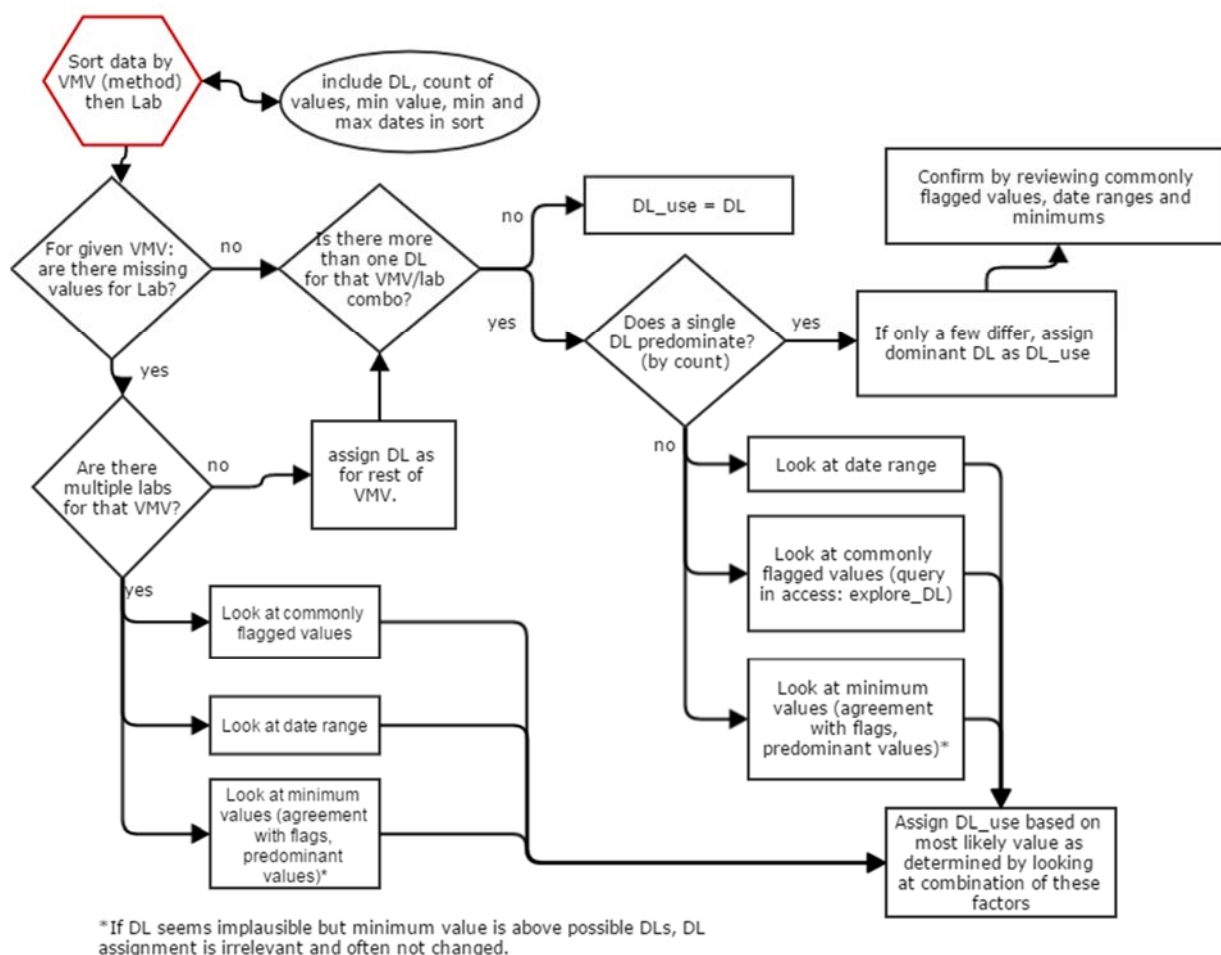
35. Total number of school age children in your household (6 to 18 years) \_\_\_\_\_

36. Total number of children that are **less than 6 years** old in your household \_\_\_\_\_

## C. Flow Charts for Water Data Analysis

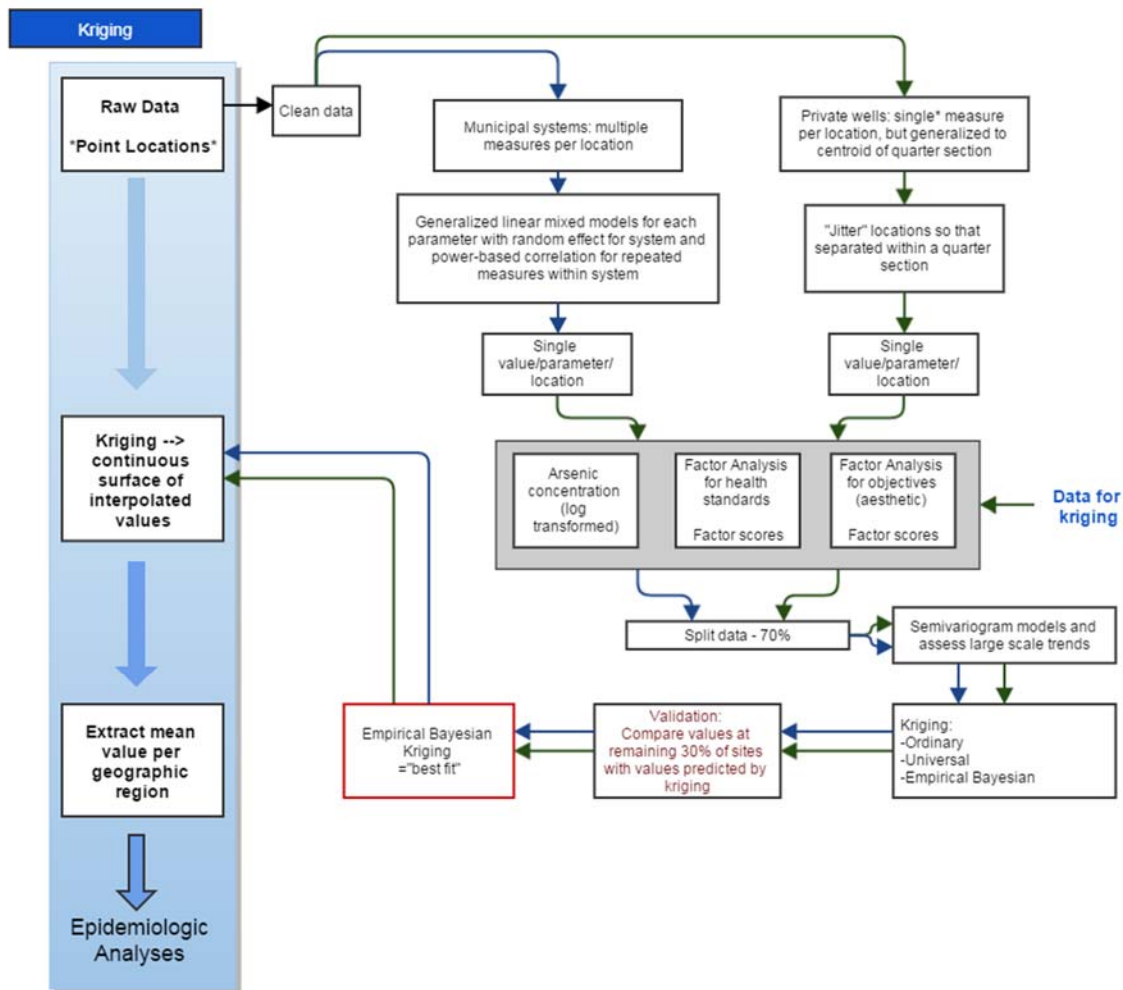
### C.1 Assign missing Detection Limit Values

Summary of process used to assign detection limit to all observations, given that detection limits changed over the study period but there were a high number of missing values for the detection limit variable. For all parameters, concentrations at or below detection limits were assigned a concentration =  $\frac{1}{2}$  the detection limit for the purposes of analysis.



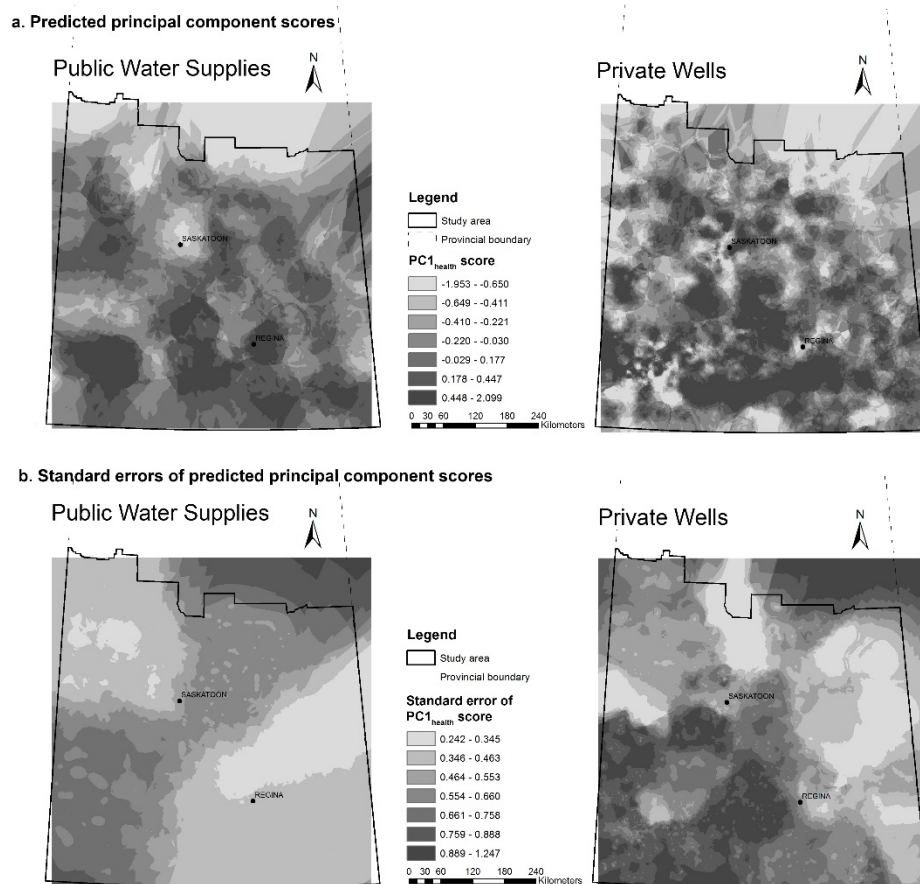
## C.2 Water Data Analysis Work Flow

This flow chart summarizes the process by which the water data was analyzed by principal components analysis and kriging. The process was applied separately to groundwater data from public water supplies and private wells.



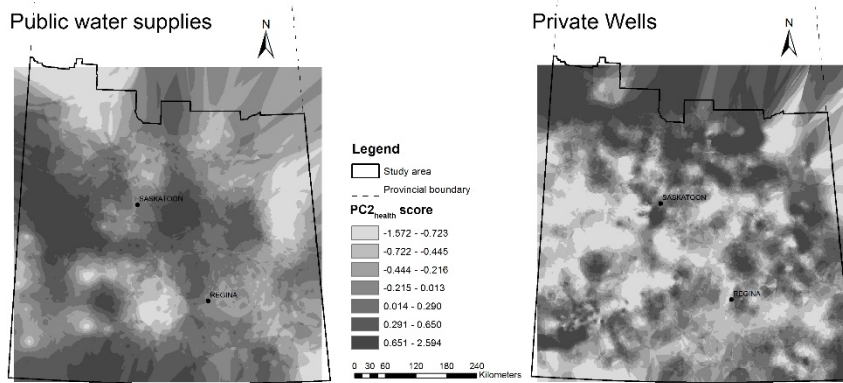
## D. Prediction Maps for Principal Component Scores

Maps showing results for final kriging predictions for principal component scores and prediction standard errors for all principal components.

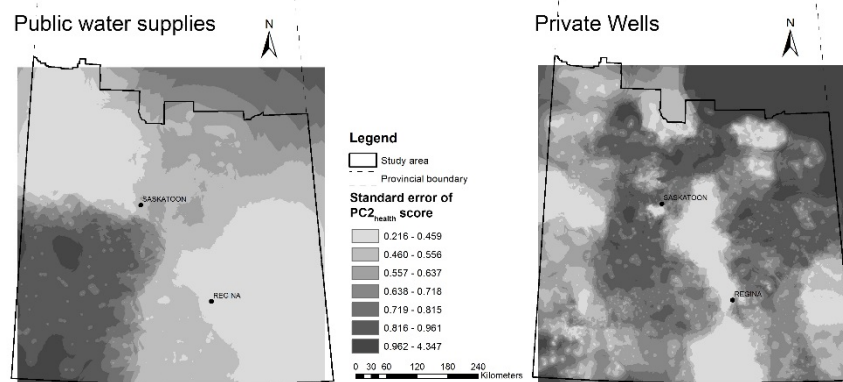


**Figure D.1** Predicted health standards first principal component scores (a) for public water supplies and private wells in study area, along with the corresponding prediction standard errors (b).

a. Predicted principal component scores

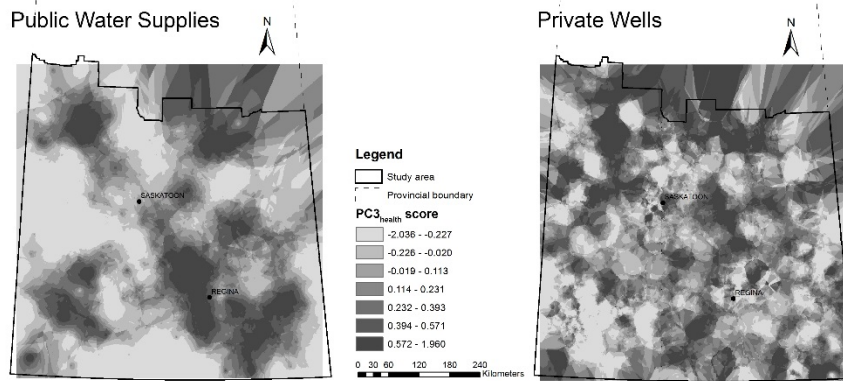


b. Standard errors of predicted principal component scores

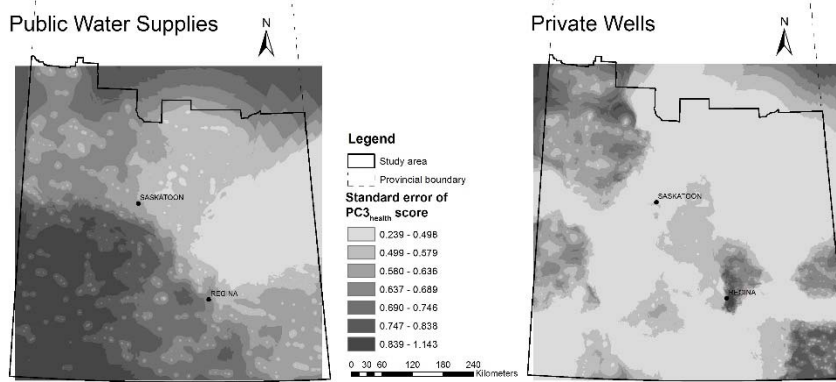


**Figure D.2** Predicted health standards second principal component scores (a) for public water supplies and private wells in study area, along with the corresponding prediction standard errors (b).

a. Predicted principal component scores

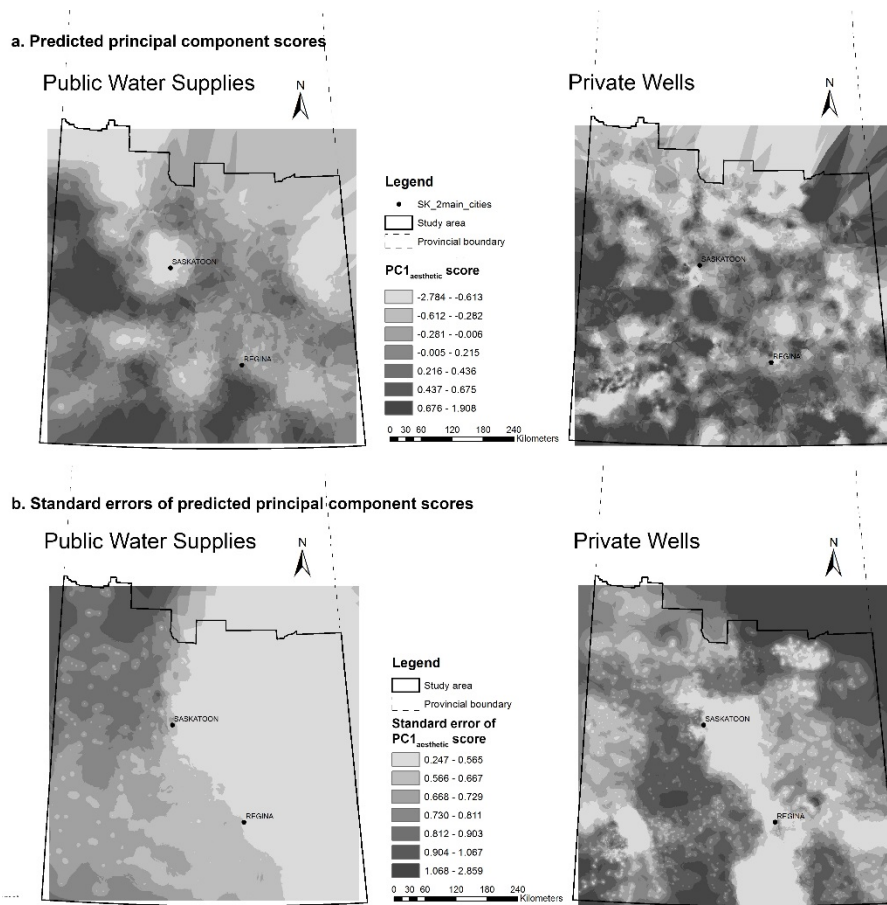


b. Standard errors of predicted principal component scores

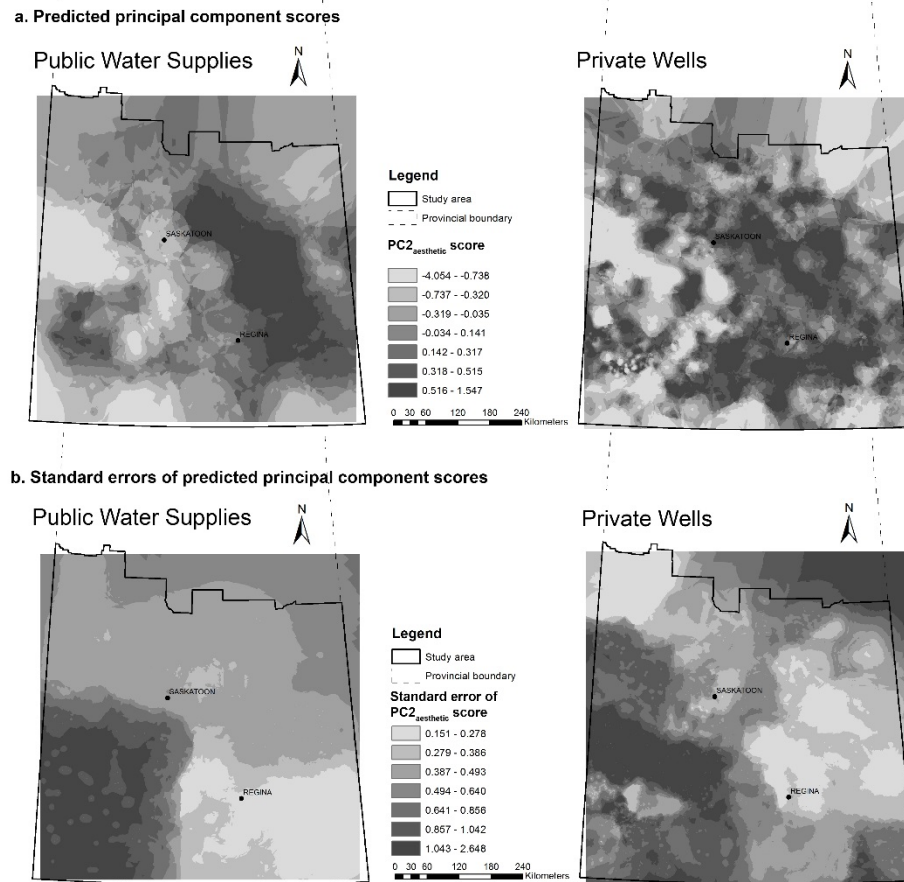


**Figure D.3** Predicted health standards third principal component scores (a) for public water supplies and private wells in study area, along with the corresponding prediction standard errors (b).



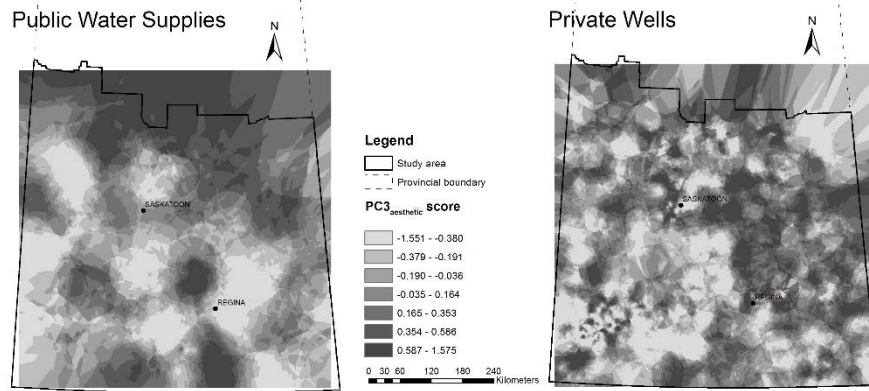


**Figure D.4** Predicted aesthetic objective first principal component scores (a) for public water supplies and private wells in study area, along with the corresponding prediction standard errors (b).

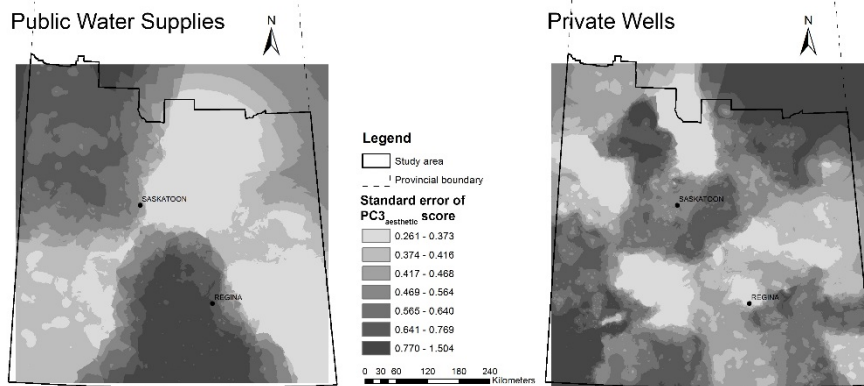


**Figure D.5** Predicted aesthetic objective second principal component scores (a) for public water supplies and private wells in study area, along with the corresponding prediction standard errors (b).

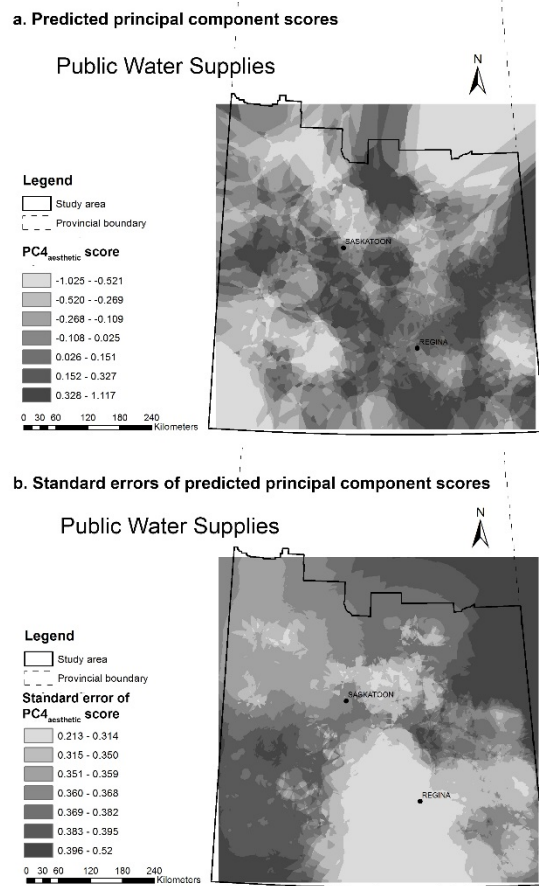
**a. Predicted principal component scores**



**b. Standard errors of predicted principal component scores**

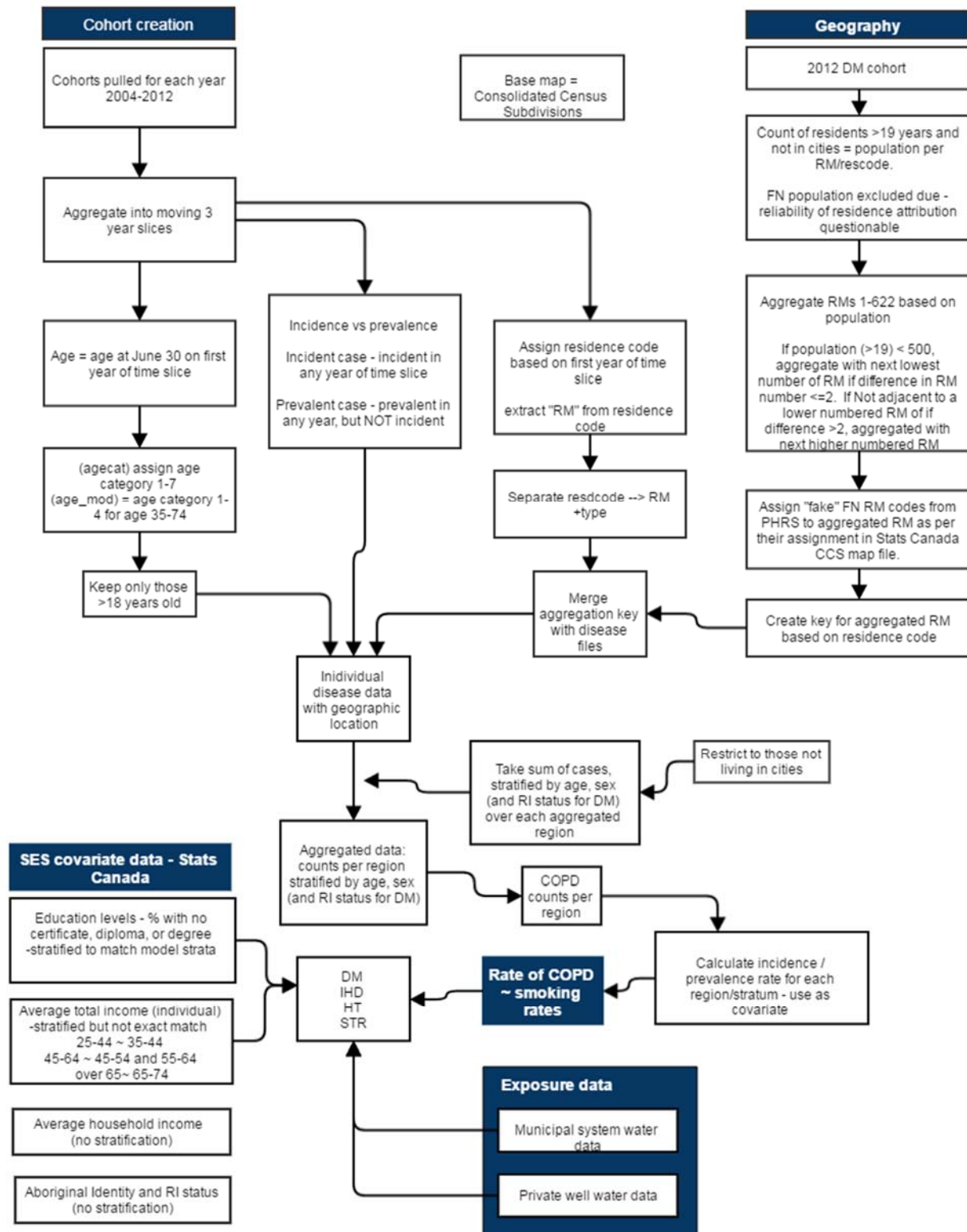


**Figure D.6** Predicted aesthetic objective third principal component scores (a) for public water supplies and private wells in study area, along with the corresponding prediction standard errors (b).



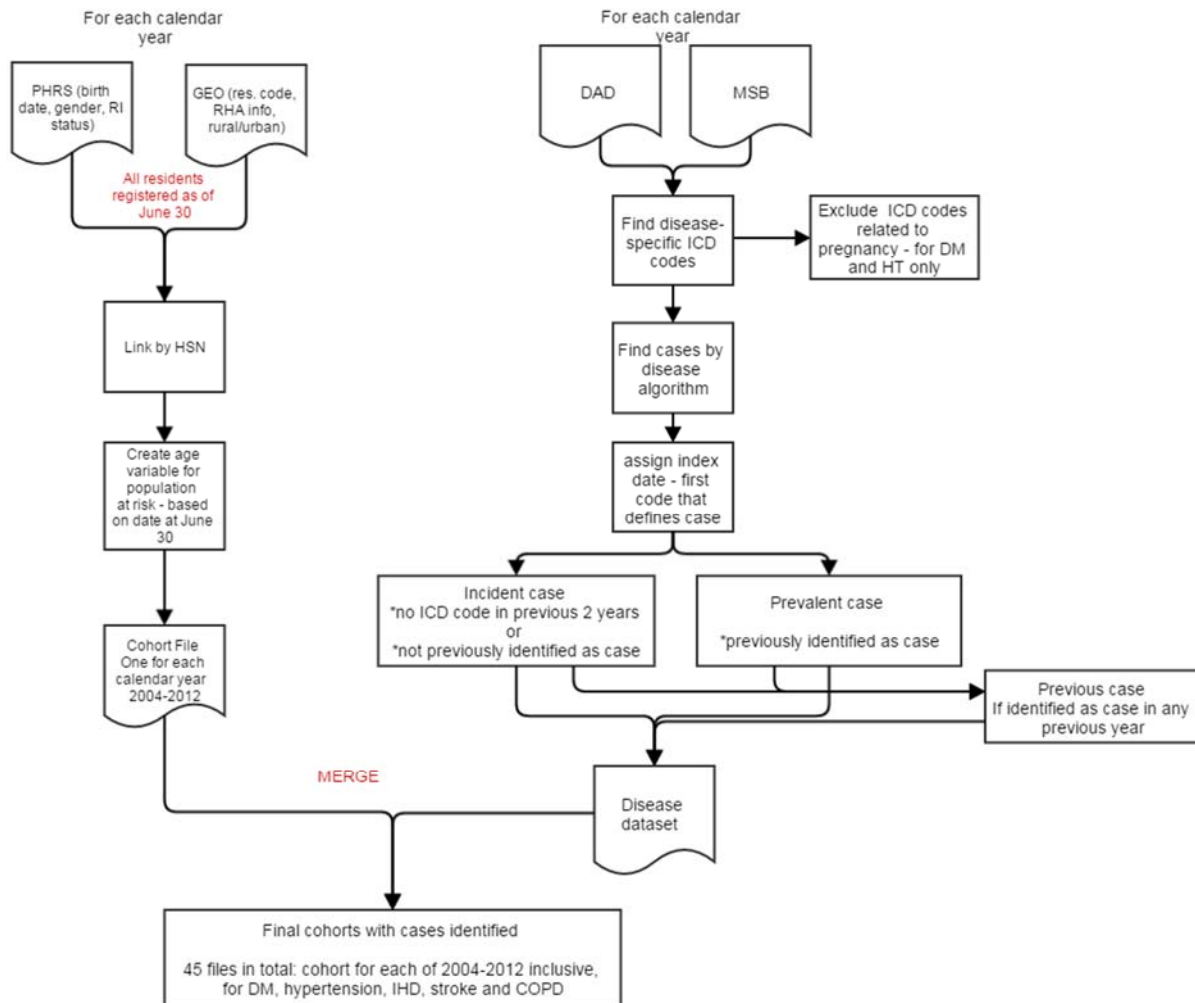
**Figure D.7** Predicted aesthetic objective fourth principal component scores (a) for public water supplies in study area, along with the corresponding prediction standard errors (b). The fourth principal component was extracted for public water supply data only.

## E. Flow chart for overall exposure-outcome analysis work flow



## F. Flow chart for cohort extraction

This flow chart summarizes the process used to extract yearly cohorts from the administrative health databases for diabetes [DM], hypertension [HT], ischemic heart disease [IHD], stroke, and chronic obstructive pulmonary disease (COPD).



PHRS = Person Registry System, GEO = geographic information linked to PHRS, DAD = hospital Discharge Abstract Database, MSB = Medical Services Branch (physician billing).

## **G. Researcher documentation for disease data management and data set creation**

The following documentation was developed to summarize data management and data set creation for the analysis of associations between water quality and diabetes and cardiovascular disease at the Health Quality Council data lab.

### **Purpose**

Main Purpose:

#1: To have no errors in reports. To accomplish this from a QMA (Quality Measurement & Analysis) perspective, we need:

- To have clear documentation for a code reviewer to take the analytical steps in the protocol and ascertain if they have been done in the programming steps.
- A communication tool between researchers and analysts to make sure all criteria for the analytical steps are covered (to achieve perfection in #1)

Other Purposes:

- A documentation tool in general for all criteria and decisions made for the analytical protocol. Therefore, it is a working document, and likely new versions for the updates would be created if indicator definitions or criteria change.
- A document archive for tracking as well as to refresh an analyst's memory a period of time later after the project completed.

### **Technical Notes**

This will be written and reorganized to match with the best of our ability to the way the analysts create programming steps, but still keeping indicator criteria logically grouped together (researcher's logic). **In the analysts' documentation**, these steps WILL likely NOT be in this same order, as the file(s) will be created keeping program efficiency in mind (analyst logic).

This document is from the researcher logic perspective, so while points will be split up by the analyst to accommodate coding, the logic will likely be linear/hierarchical in the sense of large to small – starting off with the larger concepts/file and whittling down to small, while keeping the logic of the denominators/numerators intact. But, researchers feel free to write out the steps as they make sense to you.

**1,2,3 ordering specifies sequential order, A,B,C specifies a task, not order dependent.**

The steps/numbering/lettering exist to

- Match to the analyst piece of documentation that supplements this researcher piece.
- The steps must be commented in the code as well, to be able to easily search for the criteria, and easily match the criteria in the code to both the researcher and analyst documentation. If not commented in the code, it can be difficult for and external reviewer to find the criteria of interest, and take a lot more time to review as well.





				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
Denom_sk_xx (create separate file for each calendar year 2004 – 2012)				
1	a	From PHRS _PERS_INFO file variables for each fiscal year (03/04 through 12/13) create file:  Include persons registered on June 30 of each year  i. keep all unique key_hsns, and vars for birth month and birth year, Registered Indian status, and sex	<ul style="list-style-type: none"><li>• Demographic data from Calendar Years 2004-2012 to be included in final analysis (9 cohorts)</li><li>• Persons registered on June 30 of any given year will be eligible for inclusion in cohort</li><li>• Data to be extracted for 03/04-12/13 fiscal years , and re-sorted into calendar years 2002-2012 (11 cohorts)</li><li>• 2002 and 2003 included in disease data only to provide 2 year latent period to distinguish incident vs prevalent cases</li></ul>	
	b	From PHRS_PERS_RESC file, for each fiscal year (01/02 through 12/13) create file:  Include persons registered on June 30 of each year  i. keep all unique key_hsns, and vars for RHA, rural/urban, residence code and RHA_code1-3 and RHA_area_share1-3,		
2	a	Link PHRS_PERS_RESC information to PHRS PERS_INFO data, and create cohort files for each calendar year 2002-2012	2002 and 2003 included to provide 2 years latency to determine incident vs. prevalent cases	

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
3		For each calendar year file, create age variable based on age at June 30 of that year.	Age based on birth month and year, assuming day 15 as birth day for all.	
<b>DM_cases_xx</b> (create separate file for each calendar year 2004 – 2012)				
1	a	<p>For each calendar year 2002-2012:</p> <p>In DAD (HOSP_F_YYYY) and MSB (MSB_Billing) find all instances of diabetes (type II) codes:</p> <ul style="list-style-type: none"> <li>i. In DAD: ICD-10: E11-E14 and ICD-9: 250</li> <li>ii. IN MSB: ICD-9: 250</li> </ul> <p>Keep HSN, date</p>	<p>Analysis to be completed for cases of diabetes from 2004-2012 but cases from 2002-2003 need to be ascertained to rule out previous visits for cases identified in 2004-2005 to distinguish incident vs. prevalent cases.</p> <p>For each calendar year, find codes for that year + 2 year run-in period (to allow for definition of prevalent cases)</p>	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
	b	<p>In DAD (HOSP_F_YYYY) and MSB find all instances of coding for gestational diabetes or pregnancy with following codes</p> <p>i. In DAD: ICD-10: O1, O21-95, O98, O99, Z37 and ICD-9: 641-676, V27</p> <p>ii. In MSB: ICD-9: 641-676, V27 Keep HSN, date</p>	<p>Used to rule out gestational diabetes.</p> <p>Codes will be required for 2001-2013 (2001 codes required to determine if codes found in latency period 2002-2003 are pregnancy related for distinguishing incident vs. prevalent cases; 2013 codes required to rule out gestational diabetes in 2012 cases). For each calendar year, find codes for that year + 3 year previous (to allow exclusion of pregnancy related diabetes for codes found in run in period for definition of prevalent cases) + 1 year following.</p> <p>*These codes can also be used to rule out pregnancy related hypertension codes</p>	
2	a	Exclude diabetes codes where a pregnancy related code occurs within 120 days before or 180 days after an instance of a relevant ICD code for a given HSN	Exclude gestational diabetes cases.	
3	a	If one DAD code or two MSB codes occur within a two year period, flag as diabetes case.	Case Definition: one hospitalization or two physician visits within a two year period (excluding pregnancy related codes).	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
4	a	Create index_date variable for cases (first instance of code used to define case as incident or prevalent).		
5	a	Check if HSN was flagged as case in any previous year for DAD or MSB diabetes related code in 2 years prior to index_date: <ul style="list-style-type: none"> <li>i. If not a previous case then flag as incident case</li> <li>ii. If previously identified as case, then flag as prevalent case</li> </ul>	Distinguish incident vs. prevalent cases.	
6	a	Keep HSN, index_date, prevalent, incident, previous.		
7	a	Merge diabetes case datasets with cohort files created in first step (for each calendar year).		
8	a	Create age variable for diabetes cases: <ul style="list-style-type: none"> <li>i. In year case first identified: age=age at index date</li> <li>ii. If previous case: age = age at index date + difference between current year and year of index date</li> </ul>	Age based on birth month and year, assuming day 15 as birth day for all.	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
9	a	Aggregate data into moving three year intervals i.e. 2004-2006 (inclusive), 2005-2007, 2006-2008, 2007-2009, 2008-2010, 2009-2011, 2010-2012*  i. Evaluate case counts in each rolling period and evaluate if level of aggregation sufficient to avoid small counts	To smooth variability in counts and reduce problems with small counts/cell	
		*Aggregation and analysis same for all diseases; see “Aggregation” section for steps		
<b>HT_cases _xx</b> (create separate file for each calendar year 2004 – 2012)				
1	a	For each calendar year 2002-2012:  In DAD and MSB find all instances of hypertension related codes:  i. In DAD: ICD-10: I10, I11, I12, I13, I15 and ICD-9: 401, 402, 403, 404, 405  ii. In MSB: ICD-9: 401, 402, 403, 404, 405 Keep HSN, date	Analysis to be completed for cases of hypertension from 2004-2012 but cases from 2002-2003 need to be ascertained to rule out previous visits for cases identified in 2004-2005 to distinguish incident vs. prevalent cases.  For each calendar year, find codes for that year + 2 year run-in period (to allow for definition of prevalent cases)	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
	b	<p>In DAD (HOSP_F_YYYY) and MSB find all instances of coding indicating pregnancy with following codes</p> <p>i. In DAD: ICD-10: O1, O21-95, O98, O99, Z37 and ICD-9: 641-676, V27</p> <p>ii. In MSB: ICD-9: 641-676, V27 Keep HSN, date</p>	<p>Used to rule out pregnancy-related hypertension.</p> <p>Codes will be required for 2001-2013 (2001 codes required to determine if codes found in latency period 2002-2003 are pregnancy related for distinguishing incident vs. prevalent cases; 2013 codes required to rule out pregnancy-related hypertension in 2012 cases). For each calendar year, find codes for that year + 3 year previous (to allow exclusion of pregnancy-related hypertension for cases in run in period for definition of prevalent cases) + 1 year following.</p> <p>*Can use codes from step 1b in diabetes case finding section above</p>	
2	a	Exclude instances of hypertension codes where a pregnancy related code occurs within 120 days before or 180 days after hypertension code for a given HSN	Exclude pregnancy related hypertension events	
3	a	If one DAD code or two MSB codes within a two year period for an individual then flag as a hypertension case.	Case Definition: one hospitalization or two physician visits within a two-year period.	
4	a	Create index_date variable for cases (first instance of code used to define case).		

			CHANGED OR UPDATED
STEP	CRITERIA		RATIONALE
5	a	Check for DAD or MSB hypertension related code in 2 years prior to index_date: <ul style="list-style-type: none"> <li>i. If no previous codes then flag as incident case</li> <li>ii. If a code exists in previous years, then flag as prevalent case</li> </ul>	Distinguish incident vs. prevalent cases.
	b	If HSN was flagged as case in any previous year, flag as previous case.	Qualifies as prevalent case in analysis, but must calculate age as increment from age at index case; code as separate variable from prevalent cases
6	a	Keep HSN, index_date, prevalent, incident, previous.	
7	a	Merge hypertension case datasets with cohort files created in first step (for each calendar year).	
8	a	Create age variable for hypertension cases: <ul style="list-style-type: none"> <li>i. In year case first identified: age=age at index date</li> <li>ii. If previous case: age = age at index date + difference between current year and year of index date</li> </ul>	Age based on birth month and year, assuming day 15 as birth day for all.

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
9	a	Aggregate data into moving three year intervals i.e. 2004-2006 (inclusive), 2005-2007, 2006-2008, 2007-2009, 2008-2010, 2009-2011, 2010-2012*  i. Evaluate case counts in each rolling period and evaluate if level of aggregation sufficient to avoid small counts	To smooth variability in counts and reduce problems with small counts/cell	
		*Aggregation and analysis same for all diseases; see “Aggregation” section for steps		
IHD_cases_xx (create separate file for each calendar year 2004 – 2012)				



		CHANGED OR UPDATED	
STEP	CRITERIA	RATIONALE	
1	<p>a For each calendar year 2002-2012:</p> <p>In DAD and MSB find all instances of IHD related codes:</p> <p>i. In DAD: ICD-10: I20, I21, I22, I23, I24, I25 and ICD-9: 410, 411, 412, 413, 414</p> <p>Procedure codes:</p> <p>a. CCI: 1.IJ.50, 1.IJ.57.GQ, 1.IJ.54, 1.IJ.76</p> <p>b. CCP: 48.02, 48.03, 48.11-48.19</p> <p>c. ICD-9-CM<sup>1</sup>: 36.01, 36.02, 36.05, 36.10-36.19</p> <p>ii. In MSB: ICD-9: 410, 411, 412, 413, 414 Keep HSN, date</p>	<p>Analysis to be completed for cases of ischemic heart disease from 2004-2012 but cases from 2002-2003 need to be ascertained to rule out previous visits for cases identified in 2004-2005 to distinguish incident vs. prevalent cases.</p> <p>For each calendar year, find codes for that year + 2 year run-in period (to allow for definition of prevalent cases)</p> <p>Note: procedure codes cover</p> <p>a. percutaneous coronary intervention (CCI: 1.IJ.50, 1.IJ.57.GQ, 1.IJ.54, CCP: 36.01, 36.02, 36.05, and ICD-9-CM<sup>1</sup>: 36.01, 36.02, 36.05)</p> <p>b. Coronary artery bypass graft (CCI: 1.IJ.76, CCP: 36.10-36.19, and ICD-9-CM<sup>1</sup>: 36.10-36.19)</p> <p><sup>1</sup> ICD-9-CM: Clinical Modification codes are referenced in a paper (Robataille et al 2013) but if they are not available we can ignore (CCDSS definition for IHD appears to only includes ICD-10 and ICD-9 codes with no procedure codes)</p>	

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
2	a	If one DAD code or two MSB codes within a one year period for an individual then flag as IHD case.	Case Definition: one hospitalization or two physician visits within a one year period.	
3	a	Create index_date variable for cases (first instance of code used to define case); keep HSN, index_date, prevalent, incident.		
4	a	Check for DAD or MSB IHD-related code in 2 years prior to index date:  i. If no previous codes then flag as incident case  ii. If a code exists in previous years then flag as prevalent case	Distinguish incident from prevalent cases	
	b	If HSN was flagged as case in any previous year, flag as previous case.	Qualifies as prevalent case in analysis, but must calculate age as increment from age at index case; code as separate variable from prevalent cases	
5	a	Keep HSN, index_date, prevalent, incident, previous.		
6	a	Merge IHD case datasets with cohort files created in first step for each calendar year.		

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
7	a	Create age variable for IHD cases: i. In year case first identified: age=age at index date ii. If previous case: age = age at index date + difference between current year and year of index date	Age based on birth month and year, assuming day 15 as birth day for all.	
8	a	Aggregate data into moving three year intervals i.e. 2004-2006 (inclusive), 2005-2007, 2006-2008, 2007-2009, 2008-2010, 2009-2011, 2010-2012* i. Evaluate case counts in each rolling period and evaluate if level of aggregation sufficient to avoid small counts	To smooth variability in counts and reduce problems with small counts/cell	
		*Aggregation and analysis same for all diseases; see “Aggregation” section for steps		
<b>STR_cases_xx</b> (create separate file for each calendar year 2004 – 2012)				

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
1	a	<p>For each calendar year 2002-2012:</p> <p>In DAD and MSB find all instances of stroke/TIA related codes:</p> <ul style="list-style-type: none"> <li>i. In DAD: ICD-10: I60, I61, I63, I64, H34.1, G45 – excluding I63 and G45 and ICD-9: 362.3, 430, 431, 434, 436, 435</li> <li>ii. In MSB: ICD-9: 362.3, 430, 431, 434, 436, 435</li> </ul> <p>Keep HSN, date</p>	<p>Analysis to be completed for cases of stroke/transient ischemic attack (TIA) from 2004-2012 but cases from 2002-2003 need to be ascertained to rule out previous visits for cases identified in 2004-2005 to distinguish incident vs. prevalent cases.</p> <p>For each calendar year, find codes for that year + 2 year run-in period (to allow for definition of prevalent cases)</p> <p>Exclusions: ICD-10:I63.6 (cerebral infarction due to central venous thrombosis) and G45.4 (transient global amnesia) – only available as 3 digit code in CIHI so used I64 and G45</p>	
2	a	If one DAD code or two MSB codes within a one year period for an individual then flag as stroke case.	Case Definition: one hospitalization or two physician visits within a one year period.	
3	a	Create index_date variable for cases (first instance of code used to define case); keep HSN, index_date, prevalent, incident.		

			CHANGED OR UPDATED
STEP	CRITERIA		RATIONALE
4	a	Check for DAD or MSB stroke-related code in 2 years prior to year of index_date: <ul style="list-style-type: none"> <li>i. If no previous codes then flag as incident case</li> <li>ii. If a code exists in previous years then flag as prevalent case.</li> </ul>	Distinguish incident from prevalent cases
	b	If HSN was flagged as case in any previous year, flag as previous case.	Qualifies as prevalent case in analysis, but must calculate age as increment from age at index case; code as separate variable from prevalent cases
5	a	Keep HSN, index_date, prevalent, incident, previous.	
6	a	Merge stroke case datasets with cohort files created in first step for each calendar year.	
7	a	Create age variable for stroke cases: <ul style="list-style-type: none"> <li>i. In year case first identified: age=age at index date</li> <li>ii. If previous case: age = age at index date + difference between current year and year of index date</li> </ul>	Age based on birth month and year, assuming day 15 as birth day for all.

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
8	a	Aggregate data into moving three year intervals i.e. 2004-2006 (inclusive), 2005-2007, 2006-2008, 2007-2009, 2008-2010, 2009-2011, 2010-2012*  i. Evaluate case counts in each rolling period and evaluate if level of aggregation sufficient to avoid small counts	To smooth variability in counts and reduce problems with small counts/cell  Note: for case counts, prevalent cases defined as either prevalent or previous case	
		*Aggregation and analysis same for all diseases; see "Aggregation" section for steps		
<b>COPD_cases_xx</b> (create separate file for each calendar year 2004 – 2012)				
1	a	For each calendar year 2004-2012:  In DAD and MSB find all instances of COPD related codes:  i. In DAD: ICD-10: J41, J42, J43, J44 and ICD-9: 491, 492, 496  ii. In MSB: ICD-9: 491, 492, 496 Keep HSN, date	Find COPD cases (Proxy for smoking status, to be used as confounder in models with each of the other diseases)  No run-in period required since no distinction between incident and prevalent required.	
2	a	If one DAD code or one MSB code within a one year period then flag as COPD case.	Case definition one hospitalization or one physician visit within a one year period.	

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
3	a	Create index_date variable for cases (first instance of code used to define case); keep HSN, index_date, prevalent, incident.		
4	a	If HSN was flagged as case in any previous year, flag as COPD case.	No need to distinguish incident from prevalent, but once identified as a case, considered a case for rest of study period.	
5	a	Keep HSN, index_date, COPD_case		
6	a	Merge COPD case datasets with cohort files created in first step (for each calendar year).		
7	a	Create age variable for COPD cases: <ul style="list-style-type: none"> <li>i. In year case first identified: age=age at index date</li> <li>ii. If previous case: age = age at index date + difference between current year and year of index date</li> </ul>	Age based on birth month and year, assuming day 15 as birth day for all.	
8	a	Aggregate data into moving three year intervals i.e. 2004-2006 (inclusive), 2005-2007, 2006-2008, 2007-2009, 2008-2010, 2009-2011, 2010-2012* <ul style="list-style-type: none"> <li>i. Evaluate case counts in each rolling period and evaluate if level of aggregation sufficient to avoid small counts</li> </ul>	To smooth variability in counts and reduce problems with small counts/cell	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
		<p>*Aggregation same for all diseases; see “Aggregation” section for steps</p> <p>*after aggregation, COPD stratum specific rates for each area calculated see “COPD RATES” for those steps.</p>		
<b>Aggregation into 3-year time slices</b>				
		Merge_disease_agDDDD, Aggr_disease_agDDDD	<p>Disease = DM, HT, IHD, STR, COPD</p> <p>DDDD = year</p>	
1		Prepare data for aggregate data into moving three year intervals i.e. 2004-2006 (inclusive), 2005-2007, 2006-2008, 2007-2009, 2008-2010, 2009-2011, 2010-2012	<p>*done in SAS program: macro recode</p> <p>creates: <i>disease.recode_disease_cyDDDD</i></p>	
	a	<p>recode variable names for ease of use –</p> <ul style="list-style-type: none"> <li>i. sex m = 1, F=0</li> <li>ii. FN = nav_stat_flag</li> </ul>		



				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
	b	Append cy2004-cy2012 to variable names to make unique prior to merge (age, sex, fn, rescode, incid, prev, age_index, death, index_dt.	Appending year to variable names allows differentiation in aggregated files	
2	a	Aggregate into 3-year time slices to get 7 files – individually code aggregation: sorted and merged on key_hsn.  i. Named according to first year of 3 year time slice (e.g., _ag2004 includes 2004-2006)  ii. Create seven 3-year time slices, DDDD 2004 through 2012	*Done in SAS program: aggr_disease  (separate program for each disease in folder “aggregation by year files”)  Creates files:  disease. disease.merge_disease_cyDDDD	
3	a	In same data step as merge, recode aggregated files by  appending yr1, yr2, and yr3 in place of corresponding cy2004,...,cy2012 by aggregation year.  (e.g., In merge_disease_ag2004, cy2004 → yr1, cy2005 → yr2, cy2006 → yr3	Create dentifier for which year in time slice now consistent for all files so easier to recode in macros	

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
3		Recode aggregated 3-year time slice files	Steps 3 – 4 are done in SAS program: recode postag	
	a	Age assigned as age at year 1	Ages to be categorized according to age at year 1 of 3-year slice	
	b	Use rescodeyr1- Substring to get: <ul style="list-style-type: none"> <li>i. RM portion (first 3 numbers)</li> <li>ii. residence type (last 2)</li> </ul>	Since exposure based on place of residence, use place of residence at start of time slide (given latent period for chronic disease development, want to know where resident lived in time prior to study period. Also, if moved during 3 year interval more interested in where lived in more distant past)	
	c	Code incidence/prevalence <ul style="list-style-type: none"> <li>i. If incident case in any of yr1, yr2, or yr3 coded as incident</li> <li>ii. If not incid, but prevalent in any year coded as prevalent case</li> </ul>		
	d	All other var assigned yr1 value as well (are same across years, eg. Sex, fn status, index_st_age_index)		
4	a	Code rural vs not rural <ul style="list-style-type: none"> <li>i. If residency type 20-29 then rural=0 else rural=1</li> </ul>	Residence types in the 20s correspond to cities. Since villages and towns would still be considered rural/remote water supplies for our purposes these are retained for analysis using public water supply exposure data	

			CHANGED OR UPDATED
STEP	CRITERIA	RATIONALE	
	b Code age categories 1=19-34, 2=35-44, 3=45-54, 4=55-64, 5=65-74, 6=75-84 and 7=85+ i. Delete if age category missing	These are full range of ages that might be of interest. (in future step, limited to 35-74 for analysis).  Creates files: <i>disease. disease.aggr_disease_cyDDDD</i>	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
5	a	<p>Based on RMs and associated CCS, create geographic areas “large enough” to mitigate small cell counts (will not avoid small cell counts, but will prevent too many zeroes)</p> <p>i. Using population by residence code from cohort in 2012 in over 18 population, combine RMs with pop (&gt;18) under 500 with adjacent RM.</p> <p>a. If RM(X) is adjacent to RM(X-1) combine with RM(X-1)</p> <p>b. Else if no RM numbered X-1 but RM(X-2) is adjacent, combine with RM(X-2)</p> <p>c. If not adjacent to either of those, combine with next highest number adjacent RM</p> <p>d. If any of the above are combined with others, RM(X) gets added to combined RM</p> <p>e. FN: use census information to figure out which RM a FN logically combines with according to CSD map (which includes CCS associated with each CSD; CCS corresponds to RM number nearly exactly (off in one border of RM of Hudson Bay)</p> <p>i. If more than one reserve/FN, choose reserve with largest population base according to 2006 census</p> <p>ii. RMs <math>\geq 800</math> in north – will not be retained</p>	<p>Will not prevent small cell counts but should minimize, and more importantly prevent too many zero counts (better for models)</p> <p>CCS = Stats Can census division equivalent to RM</p> <p>Based on how RMs are numbered in province (East to West and South to North), this algorithm worked well to logically group RMs and is reasonably arbitrary (e.g. at Eastern border, next lower RM likely to be on West side of province, so go up instead). Occasionally RMs have been folded together to X-2 rule seems to nearly follow number convention (i.e. X-2 likely to be adjacent if no RM( X-1) exist but larger jumps appear to reflect E-W, N-S number jumps)</p> <p>If FN, rescode assigned to home reserve; most likely to be that with reasonably sized population on census in terms of location. Still a limitation in that FN “home reserve” may have nothing to do with actual place of residence.</p> <p>Created by hand in Excel as S:\Data\Working\new_CCS_aggr_toSAS.xlsx</p> <p>imported to sas as working.ccs_aggr_fileneu</p>	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
6	a	Merge CCS aggregation information with cohort files	<p>Sas program: aggr_by_RM_new</p> <p>(new = had to re-do based on inconsistencies between cov pop files on which first aggregation attempt based)</p> <p>Creates files:</p> <p><i>disease. disease.aggrgeo2_disease_cyDDDD</i></p>	
	b	Assign BugsID variable based on CCS aggregation (simple incremented variable to correspond with ascending CCS_aggrgation numbers since this is how OpenBUGS identifies geographic areas)	working.bugs_aggr_filenew;	
7	a	Merge each disese file file with appropriate BugsID	<p>Creates files:</p> <p><i>disease. disease.agrbugsnew_disease_agDDDD</i></p>	
	b	<p>Recode age categories for Mapping Analysis -&gt; age_mod</p> <p>1=35-44, 2=45-54, 3=55-64, 4=65-74</p>	note program also created an over55 variable – this was only for non stratified data for GeoMED presentation	

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
8	a	<p>Sum incid and prev per stratum for each region using PROC MEAN step to output count files; n(cases) is just the number of observations per stratum, equals number at risk (total population – must subtract #prevalent cases to get at risk for incidence)</p> <p>-by BugslD, sex, age_mod and fn for diabetes and copd, (for copd this is for just for calculating rates as covariate, for diabetes is the case count file)</p> <p>-by BugslD, sex and age_mod for other outcomes</p> <p>(-also took sum of cases, but dropped later (not used), for some reason took means of bugslD or aggCCS but not necessary)</p>	<p>SAS program: macro summarize new</p> <p>Creates files:</p> <p>Dm.summFN_dm_agDDDD</p> <p>Copd.summFN_copd_agDDDD</p> <p>Ht.summ2_ht_agDDDD</p> <p>ihd.summ2_ihd_agDDDD</p> <p>str.summ2_str_agDDDD</p>	
<b>Calculate Rates and Expected Counts</b>				
1	a	<p>HT, IHD, STR:</p> <p>Sum number of incident cases , prevalent cases, and population at risk by age and sex (use PROC MEANS on <i>disease.summ2_disease_agDDDD</i>) to create <i>disease.agesextot_disease_agDDDD</i> files for each disease and time period</p>	<p>For each outcome under study:-calculate incidence and prevalence rates stratified by sex and age across all areas to get rates across study area for calculating expected number of cases (i.e. study wide rate*pop at risk = expected counts)</p> <p>-additional level of stratification by FN status for diabetes outcome (not for COPD, see treatment of COPD as covariate in separate section below)</p> <p>Program: macro calc rates</p>	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
2	a	Data step to assign stratum number to each age/sex combination in disease.agesextot_disease_agDDDD files and calculate (for each age/sex stratum )  - incidence rate (inc_risk) = number of incident cases/(total population – number prevalent cases)  - prevalence rate (prev_risk = total counts / total pop at risk	Number strata 1-8 to created merge variable	
	b	Data step to assign stratum number to each age/sex combination in disease.summ2_disease_agDDDD files	Number strata 1-8 to created merge variable	
3	a	Merge corresponding area-wide rates to disease summary files according to stratum number	Now have expected rate by each sex/age group merged to data for each region	
4	a	Calculate expected counts  -at risk for prevalence = atrisk (total population/strata)  -calculated at risk for incidence (atrisk_i) = atrisk-number of prevalent cases		
	b	Calculate expected counts for incident and prevalent cases by multiplying inc_risk and prev_risk by atrisk_i and atrisk respectively →exp_incid and exp_prev	Expected rate by region * population at risk = expected count by region for each stratum for modeling  →disease.obsexp_disease_agDDDD	

		STEP	CRITERIA	RATIONALE	CHANGED OR UPDATED
5	a		<p>Data step to calculate regional rates for COPD without FN stratification from copd.obsexp_copd_agDDDD</p> <ul style="list-style-type: none"> <li>Sum of cases / at risk population in each stratum (<math>\text{num\_incid}/(\text{atrisk}-\text{num\_prev}) = \text{copd\_in\_risk}</math>; <math>\text{num\_prev}/\text{atrisk} = \text{copd\_pr\_risk}</math>) – <b>only prevalence variable used as covariate</b></li> <li>Inc_risk and inc_rate variables renames as SKcopd_inc_rate and SKcopd_pr_rate (NOT used)</li> </ul>	<p>COPD as covariate – want rate by region/age/sex as covariate. Because available, also included overall rate across regions, but this not used in this analysis - <b>only prevalence variable used as covariate</b></p> <p>→copd.regionrates_copd_agDDDD</p>	
	a	6	<p>DM</p> <p>Sum number of incident cases, prevalent cases, and population at risk by age,sex, and FN status (by PROC MEANS on dm.summfn_dm_agDDDD) to create dm.fnagesextot_dm_agDDDD files for each disease and time period</p>		
	a	7	<p>Data step to assign stratum number to each age/sex/FN combination in dm.fnagesextot_dm_agDDDD files and calculate (for each age/sex stratum )</p> <ul style="list-style-type: none"> <li>incidence rate (inc_risk) = number of incident cases/(total population – number prevalent cases)</li> <li>prevalence rate (prev_risk = total counts / total pop at risk</li> </ul>	Number strata 1-16 to create merge variable	



		STEP	CRITERIA	RATIONALE	CHANGED OR UPDATED
	b		Data step to assign stratum number to each age/sex/FN combination in dm.summfn_dm_agDDDD files	Number strata 1-16 to create merge variable	
8	a		Merge corresponding area-wide rates to disease summary files according to stratum number	Now have expected rate by each sex/age group merged to data for each region	
9	a		Calculate expected counts -at risk for prevalence = atrisk (total population/strata)  -calculated at risk for incidence (atrisk_i) = atrisk - number of prevalent cases	Expected rate by region * population at risk = expected count by region for each stratum for modeling  → dm.fnobsexp_dm_agDDDD	
	b		Calculate expected counts for incident and prevalent cases by multiplying inc_risk and prev_risk by atrisk_i and atrisk respectively → exp_incid and exp_prev		

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
10	a	<p>Data step to calculate regional rates for COPD <b>with FN stratification</b> from copd.fnobsexp_copd_agDDDD</p> <ul style="list-style-type: none"> <li>Sum of cases / at risk population in each stratum (<math>\text{num\_incid}/(\text{atrisk}-\text{pnum\_prev}) = \text{copd\_in\_risk}</math>; <math>\text{num\_prev}/\text{atrisk} = \text{copd\_pr risk}</math>) – <b>only prevalence variable used as covariate</b></li> </ul> <p>Inc_risk and inc_rate variables renames as SKcopd_inc_rate and SKcopd_inc_rate (NOT used)</p>	<p>COPD as covariate – want rate by region/age/sex?FN status as covariate. Because available, also included overall rate across regions, but this not used in this analysis - <b>only prevalence variable used as covariate</b></p> <p>→copd.fnregionrates_copd_agDDDD</p>	
<b>Merge Covariates: COPD rates, Education, Income, COPD rates</b>				
1	a	<p>Merge COPD rates:</p> <ul style="list-style-type: none"> <li>copd.regionrates_copd_agDDDD</li> </ul> <p>With each of the outcome files:</p> <ul style="list-style-type: none"> <li>ht.obsexp_ht_agDDDD</li> <li>ihd.obsexp_ihd_agDDDD</li> <li>str.obsexp_str_agDDDD</li> </ul> <p>after sorting by merge variables : bugslD, sex, age</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>ht.merge1_ht_agDDDD</li> <li>ihd.merge1_ihd_agDDDD</li> <li>str.merge1_str_agDDDD</li> </ul>	<p>Do steps for non FN stratified outcomes using macros, repeated (see below) in separate macros for FN stratified data (diabetes)</p> <p>SAS program: macro merge_census</p>	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
2	a	<p>Merge education covariates:</p> <ul style="list-style-type: none"> <li>• Census.educ_merge</li> </ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"> <li>• ht.merge1_ht_agDDDD</li> <li>• ihd.merge1_ihd_agDDDD</li> <li>• str.merge1_str_agDDDD</li> </ul> <p>by merge variables : bugsID, sex, age</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>• ht.merge2_ht_agDDDD</li> <li>• ihd.merge2_ihd_agDDDD</li> <li>• str.merge2_str_agDDDD</li> </ul>	<p>Education data file created external to HQC using Census of Canada 2006 data. By regions, stratified by sex and age (age categories correspond exactly to cohort age groups)</p>	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
3	a	<p>Merge individual income covariates</p> <ul style="list-style-type: none"> <li>• Census.income_merge</li> </ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"> <li>• ht.merge2_ht_agDDDD</li> <li>• ihd.merge2_ihd_agDDDD</li> <li>• str.merge2_str_agDDDD</li> </ul> <p>by merge variables : bugslD, sex, age</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>• ht.merge3_ht_agDDDD</li> <li>• ihd.merge3_ihd_agDDDD</li> <li>• str.merge3_str_agDDDD</li> </ul>	<p>Individual income by region, stratified by sex and age but age categories not exactly match for this data. Census data for ages 25-44 assigned to cohort age group 35-44, census data for 45-64 assigned to cohort age groups 45-54 and 55-64, and census data for over 65 assigned to cohort age group 65-74.</p> <p>This variable was then centered and scaled by a factor of 1000 to improve interpretability.</p>	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
4	a	<p>Merge household income covariates</p> <ul style="list-style-type: none"> <li>• Census.hshld_merge</li> </ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"> <li>• ht.merge3_ht_agDDDD</li> <li>• ihd.merge3_ihd_agDDDD</li> <li>• str.merge3_str_agDDDD</li> </ul> <p>by merge variables : bugsID</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>• ht.merge4_ht_agDDDD</li> <li>• ihd.merge4_ihd_agDDDD</li> <li>• str.merge4_str_agDDDD</li> </ul>	Household income is by region only. Added as a “just in case” covariate but not used in analysis	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
5	a	<p>Merge aboriginal identity covariates</p> <ul style="list-style-type: none"> <li>• Census.aborig_ident</li> </ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"> <li>• ht.merge4_ht_agDDDD</li> <li>• ihd.merge4_ihd_agDDDD</li> <li>• str.merge4_str_agDDDD</li> </ul> <p>by merge variables : bugslD</p> <p>Recode a series of extra variables by hand with SK averages and whether value above or below SK average (for each covariate)</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>• ht.merge5_ht_agDDDD</li> <li>• ihd.merge5_ihd_agDDDD</li> <li>• str.merge5_str_agDDDD</li> </ul>	<p>Aboriginal identity variable from census data also available at region level only. Another “just in case” variable not used in analysis.</p> <p>Variables coded for over/under SK averages, also not used.</p>	

		STEP	CRITERIA	RATIONALE	CHANGED OR UPDATED
6	a		<p>Repeat but for FN stratified diabetes data</p> <p>Merge COPD rates:</p> <ul style="list-style-type: none"> <li>• copd.fnregionrates_copd_agDDDD</li> </ul> <p>With each of the outcome files:</p> <ul style="list-style-type: none"> <li>• dm.obsexp_dm_agDDDD</li> </ul> <p>after sorting by merge variables : bugslD, sex, FN, age category</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>• dm.FNmerge1_dm_agDDDD</li> </ul>	<p>Do steps for FN stratified diabetes outcomes using similar set of macros (only first is different to account for FN stratified COPD data; census covariates not FN stratified)</p> <p>Still in SAS program: macro merge census</p>	
	a	7	<p>Merge education covariates:</p> <ul style="list-style-type: none"> <li>• Census.educ_merge</li> </ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"> <li>• dm.FNmerge1_dm_agDDDD</li> </ul> <p>by merge variables : bugslD, sex, age</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>• dm. FNmerge2_dm_agDDDD</li> </ul>	<p>Merge all census covariates to FN stratified files</p> <p>Education data file created external to HQC using Census of Canada 2006 data. By regions, stratified by sex and age (age categories correspond exactly to cohort age groups)</p> <p>*No FN stratification in census data, however</p>	

				CHANGED OR UPDATED
STEP		CRITERIA	RATIONALE	
8		<p>Merge total individual income covariates:</p> <ul style="list-style-type: none"><li>Census.income_merge</li></ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"><li>dm.FNmerge2_dm_agDDDD</li></ul> <p>by merge variables : bugslD, sex, age</p> <p>Creates:</p> <ul style="list-style-type: none"><li>dm.FNmerge3_dm_agDDDD</li></ul>	<p>Individual income by region, stratified by sex and age but age categories not exactly match for this data. Census data for ages 25-44 assigned to cohort age group 35-44, census data for 45-64 assigned to cohort age groups 45-54 and 55-64, and census data for over 65 assigned to cohort age group 65-74.</p> <p>This variable was then centered and scaled by a factor of 1000 to improve interpretability.</p>	
9	a	<p>Merge household income covariate:</p> <ul style="list-style-type: none"><li>Census.hshld_merge</li></ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"><li>Dm.FNmerge3_dm_agDDDD</li></ul> <p>by merge variables : bugslD, sex, age</p> <p>Creates:</p> <ul style="list-style-type: none"><li>dm.FNmerge4_dm_agDDDD</li></ul>	<p>Household income is by region only. Added as a “just in case” covariate but not used in analysis</p>	



STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
10	a	<p>Merge household income covariate:</p> <ul style="list-style-type: none"> <li>Census.hshld_merge</li> </ul> <p>With each of the previously created files:</p> <ul style="list-style-type: none"> <li>dm.FNmerge4_dm_agDDDD</li> </ul> <p>by merge variables : bugslD, sex, age</p> <p>Recode a series of extra variables by hand with SK averages and whether value above or below SK average (for each covariate)</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>dm.FNmerge5_dm_agDDDD</li> </ul>	<p>Aboriginal identity variable from census data also available at region level only. Another “just in case” variable not used in analysis.</p> <p>Variables coded for over/under SK averages, also not used.</p>	

				CHANGED OR UPDATED
STEP	CRITERIA		RATIONALE	
Merge Water data				
1	a	<p>Import and merge public supply water data.</p> <p>water.largesys_merge → drop extraneous variables to create water.largesys_clean then merge with each of</p> <ul style="list-style-type: none"><li>dm.FNmerge5_dm_agDDDD</li><li>ht.merge5_ht_agDDDD</li><li>ihd.merge5_ihd_agDDDD</li><li>str.merge5_str_agDDDD</li></ul> <p>Creates</p> <ul style="list-style-type: none"><li>dm.FNlgesys_dm_agDDDD</li><li>ht.lgesys_ht_agDDDD</li><li>ihd.lgesys_ihd_agDDDD</li><li>str.lgesys_str_agDDDD</li></ul>	<p>SAS program: <b>import and prep water</b></p> <p>And <b>macro merge water AUG2016</b></p> <p>Separate macros for FN stratified (diabetes). Water variables are mean predicted value for each variable on a per region bases; merge only on BugsID (geographic area id variable).</p>	

		STEP	CRITERIA	RATIONALE	CHANGED OR UPDATED
	2	a	<p>Import and Merge RWQAP (private supply) water data</p> <p>-water.rwqap_mergeMAY - drop extraneous variables and convert arsenic to µg/L → water.rwqap_clean</p> <p>with</p> <ul style="list-style-type: none"> <li>• dm.FNlgesys_dm_agDDDD</li> <li>• ht.lgesys_ht_agDDDD</li> <li>• ihd.lgesys_ihd_agDDDD</li> <li>• str.lgesys_str_agDDDD</li> </ul> <p>Creates</p> <ul style="list-style-type: none"> <li>• dm.FNlallwaterAUG_dm_agDDDD</li> <li>• ht.allwaterAUG_ht_agDDDD</li> <li>• ihd.allwaterAUG_ihd_agDDDD</li> <li>• str.allwaterAUG_str_agDDDD</li> </ul>	<p>New water data was created in may due to location attribution issue with private well data. Corrected file = water.rwqap_mergeMAY</p> <p>[Note originally the new water data was merged with existing then cleaned up to create files to Export for use in OpenBUGS as .csv</p> <ul style="list-style-type: none"> <li>• R_FNwaterMAY_dm_ag2010.csv</li> <li>• R_waterMAY_ht_ag2010.csv</li> <li>• R_waterMAY_ihd_ag2010.csv</li> <li>• R_waterMAY_str_ag2010.csv</li> </ul> <p>And were manually edited to create:</p> <p>(with corrected RWQ data)</p> <ul style="list-style-type: none"> <li>• dm_2010_MAYtxt.txt</li> <li>• ht_2010_MAYtxt.txt</li> <li>• ihd_2010_MAYtxt.txt</li> <li>• str_2010_MAYtxt.txt</li> </ul> <p>Though this data was used for most of analysis, in August this was modified to create new permanent files to be used from August 2016 on.</p>	<p>May 2016</p> <p>And</p> <p>Aug 2016</p>

STEP	CRITERIA	RATIONALE	CHANGED OR UPDATED
3	<p>a Clean data for analysis</p> <p>For:</p> <ul style="list-style-type: none"> <li>dm. FNlallwaterAUG_dm_agDDDD</li> <li>ht. allwaterAUG_ht_agDDDD</li> <li>ihd. allwaterAUG_ihd_agDDDD</li> <li>str. allwaterAUG_str_agDDDD</li> </ul> <p>-delete if atrisk=0 or atrisk=. for prevalence data, or atrisk_i =0 or . for incidence data (where atrisk_i=population-#prevalent cases).</p> <p>-calculate offset variable for models (ln(expected))</p> <p>-income is in thousands – centre and scale by 1000 for interpretability</p> <p>Creates:</p> <ul style="list-style-type: none"> <li>dm. POffset_FNwater_dm_agDDDD and</li> <li>dm. lNoffset_FNwater_dm_agDDDD</li> <li>ht. offset_water_ht_agDDDD</li> <li>ihd.offset_water_ihd_agDDDD</li> <li>str.offset_water_str_agDDDD</li> </ul>	<p>Because OpenBugs does not handle missing data easily, delete any observation with at risk population =0 (or missing) because where at risk=0, expected count will be 0; need ln(expected) for passion models and this will be missing if atrisk=0.</p> <p>Create separate files for prevalence and incidence - there are 73 strata that had one person at risk and one prevalent case, therefore had 0 at risk for incidence (but not prevalence) so had to delete more rows for incidence models</p> <p>-water data had extra BugsID for north – this step also discards these extra lines.</p>	Aug 2016

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
4	a	Create categorical version of water variables from public supplies with PROC RANK on water.lgesys_clean → quintiles for water.lgesys_ranked	SAS program: macro water categories AUG  Some interesting effects found with continuous variables, but need to double check relationships with categorical variables. This was only done for 2010 models (2010-2012) for diabetes incidence) where indicated by initial results	
	b	Create categorical version of water variables from private supplies with PROC RANK on water.rwqap_clean → quintiles in water.rwqap_ranked		
5	a	Create dummy coded versions of categorical variables in water.lgesys_ranked with first quintile as reference to create water.lgesys_dummy	Now creates set of dummy coded variables for categories for use in OpenBUGS. Resulting files have water data as categorical variables and dummy coded categorical variables	
	b	Create dummy coded versions of categorical variables in water.rwqap_ranked with first quintile as reference to create water.rwqap_dummy	compared categorical and dummy variables in Proc Glimmix to confirm logic/coding correct.	

STEP		CRITERIA	RATIONALE	CHANGED OR UPDATED
6	a	<p>Merge water.lgesys_dummy with 2010 by bugsID with:</p> <ul style="list-style-type: none"> <li>• dm. POffset_FNwater_dm_ag2010</li> <li>• dm. INoffset_FNwater_dm_ag2010</li> <li>• ht. offset_water_ht_ag2010</li> <li>• ihd.offset_water_ihd_ag2010</li> <li>• str.offset_water_str_ag2010</li> </ul> <p>to create data with dummy variables:</p> <ul style="list-style-type: none"> <li>• water.DM_PRwatercatlgeAUG</li> <li>• water.DM_INwatercatlgeAUG</li> <li>• water.HT_watercatlgeAUG</li> <li>• water.IHD_watercatlgeAUG</li> <li>• water.STR_watercatlgeAUG</li> </ul>	<p>Merge full disease+water files with dummy variables – only for 2010/2010-12 because only looking in detail at this time period</p> <p>Did separately for public (lge) and private (rwq) water supplies to facilitate reformatting data for OpenBugs.</p> <p>*note, each file contains continuous water variables from both supplies, but categorized for only one type of supply</p>	

STEP	CRITERIA	RATIONALE	CHANGED OR UPDATED
	<p>b Merge water.rwqap_dummy with 2010 by bugsID with:</p> <ul style="list-style-type: none"> <li>• dm. POffset_FNwater_dm_ag2010</li> <li>• dm. INoffset_FNwater_dm_ag2010</li> <li>• ht. offset_water_ht_ag2010</li> <li>• ihd.offset_water_ihd_ag2010</li> <li>• str.offset_water_str_ag2010</li> </ul> <p>to create data with dummy variables:</p> <ul style="list-style-type: none"> <li>• water.DM_PRwatercatrwqAUG</li> <li>• water.DM_INwatercatrwqAUG</li> <li>• water.HT_watercatrwqAUG</li> <li>• water.IHD_watercatrwqAUG</li> <li>• water.STR_watercatrwqAUG</li> </ul>		Aug 2016
7	<p>a Export files to .csv format for transfer to OpenBugs for categorized large</p> <ul style="list-style-type: none"> <li>• DMPRwatercatlgeAUG.csv</li> <li>• DMINwatercatlgeAUG.csv</li> <li>• HTwatercatlgeAUG.csv</li> <li>• IHDwatercatlgeAUG.csv</li> <li>• STRwatercatlgeAUG.csv</li> </ul>		

			CHANGED OR UPDATED
STEP	CRITERIA		RATIONALE
	b	<p>Export files to .csv format for transfer to OpenBugs for categorized RWQ data</p> <ul style="list-style-type: none"> <li>• DMPRwatercat_rwqAUG.csv</li> <li>• DMINwatercat_rwqAUG.csv</li> <li>• HTwatercat_rwqAUG.csv</li> <li>• IHDwatercat_rwqAUG.csv</li> <li>• STRwatercat_rwqAUG.csv</li> </ul>	
8	a	<p>Manually edit files for OpenBugs format, deleting extraneous variables for each to create:</p> <ul style="list-style-type: none"> <li>• dmp2010_lge_catAUGt.txt</li> <li>• dmi2010_lge_catAUGt.txt</li> <li>• dmi2010_rwq_catAUGt.txt</li> <li>• ht2010_rwq_catAUGt.txt</li> <li>• ihd2010_lge_catAUGt.txt</li> <li>• str2010_lge_catAUGt.txt</li> </ul>	<p>These files are the final ones that should be used for any OpenBugs analysis</p> <p>Note process creates transposed .CSV files with same name without t suffix.</p>